Census Recensement

Canada 1986

GUIDE A L'INTENTION
DES UTILISATEURS
SUR LA QUALITÉ DES DONNÉES
DU RECENSEMENT DE 1986:

ÉCHANTILLONNAGE ET
PONDÉRATION

Canada

Données de catalogage avant publication (Canada)

Vedette principale au titre:

Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : échantillonnage et pondération

(Référence)
(Census Canada, 1986 - Recensement Canada, 1986)
Publié aussi en anglais sous le titre: User's guide to the quality of 1986 census data, sampling and weighting.
ISBN 0-660-92752-7
CS99-136F

1. Canada -- Recensement, 1986 -- Méthodologie. 2. Canada -- Recensement, 1986 -- Evaluation. I. Statistics Canada. II. Titre: User's guide to the quality of 1986 census data, sampling and weighting. III. Collection: Référence. IV. Collection: Census Canada, 1986 - Recensement Canada, 1986.

HA741.5.1986 U8314 1990   304.6'0723 C90-098028-1



Recensement
Census

Canada 1986

Référence

GUIDE À L'INTENTION DES UTILISATEURS SUR LA QUALITÉ DES DONNÉES DU RECENSEMENT DE 1986:

ÉCHANTILLONNAGE ET PONDÉRATION

Publication autorisée par le ministre de l'Industrie, des Sciences et de la Technologie

Octobre 1990

Prix: Canada, $23.00
Autres pays, $24.00

Catalogue 99-136F

ISBN-0-660-92752-7

Ottawa

This publication is available in English upon request (Catalogue No. 99-136E).

Données de catalogage avant publication (Canada)

Vedette principale au titre:

Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : échantillonnage et pondération

(Référence)
(Census Canada, 1986 - Recensement Canada, 1986)
Publié aussi en anglais sous le titre: User's guide to the quality of 1986 census data, sampling and weighting.
ISBN 0-660-92752-7
CS99-136F

1. Canada -- Recensement, 1986 -- Méthodologie. 2. Canada -- Recensement, 1986 -- Evaluation. I. Statistics Canada. II. Titre: User's guide to the quality of 1986 census data, sampling and weighting. III. Collection: Référence. IV. Collection: Census Canada, 1986 - Recensement Canada, 1986.

HA741.5.1986 U8314 1990   304.6'0723   C90-098028-1

REMERCIEMENTS

Ce document a été préparé par Stephen Rathwell de la Division des Méthodes d'Enquêtes Sociales.

L'auteur aimerait remercier Mike Bankier, Pierre Daoust, Yves Béland, Don Royce, et Dick Carter de la Sous-Division des Méthodes de Recensement et Recherches pour leur participation à l'élaboration de ce document.

L'auteur voudrait également remercier Louise Whalen et Lorna Bailie de la Division des Opérations du Recensement pour leur contribution.

Certaines parties de ce document, plus spécialement dans les trois premiers chapitres, proviennent de la publication Qualité des donnés, Série 1: Sources d'erreurs - Échantillonnage et pondération, du Recensement de 1976. Ce document avait été préparé à l'époque par Gordon Brackstone.

TABLE DES MATIÈRES

Liste des tableaux

TABLE DES MATIÈRES - fin

Liste des figures

I. INTRODUCTION

De nos jours, l'échantillonnage constitue une technique largement utilisée dans maints domaines. Nous pouvons juger de la qualité d'un produit à l'aide d'un échantillon avant d'en faire l'achat; nous nous formons une opinion des gens à partir d'échantillons de leur comportement; nos impressions sur tel pays ou telle ville sont basées sur les courts voyages que nous y avons faits. Ce sont là des exemples d'échantillonnage menant à tirer des conclusions sur le "tout" à partir de connaissances portant sur une "partie".

Scientifiquement parlant, l'échantillonnage est utilisé par les comptables, par exemple, lorsqu'ils effectuent la vérification des états financiers; dans l'industrie, pour vérifier la qualité d'articles produits en série, et par les initiateurs de sondages d'opinion et d'enquêtes pour recueillir des renseignements sur la façon de penser ou les caractéristiques d'une population donnée. On a généralement recours à l'échantillonnage pour réduire les coûts d'enquête ou pour obtenir rapidement des données, ou les deux. Dans certains cas, le mesurage peut même détruire le produit à l'étude (par exemple, l'évaluation de la durée de vie des ampoules électriques) et par conséquent, l'échantillonnage s'avère essentiel. Le désavantage de cette méthode est que les résultats obtenus à partir d'un échantillon peuvent ne pas être aussi précis que ceux recueillis auprès de la population totale. Toutefois, l'échantillonnage peut être avantageux quand la perte de précision (qui peut être minime si l'échantillon est suffisamment grand) est acceptable en fonction de l'utilisation prévue des résultats. De plus, la réduction de l'envergure d'une étude faite à partir d'un échantillon peut mener à la réduction des erreurs non dues à l'échantillonnage, ce qui compense en partie le manque de précision attribuable à l'échantillonnage.

Pour le recensement de la population de 1986, on a appliqué les méthodes d'échantillonnage de plusieurs façons. On y a eu recours pour s'assurer que la qualité du travail des représentants du recensement dans le cadre de la collecte des questionnaires respectait certaines normes; on y a eu recours pour le contrôle qualitatif du codage des réponses au cours du dépouillement dans les bureaux régionaux; on y a eu recours pour estimer le sous-dénombrement et le surdénombrement qui se sont produits pour diverses raisons; on y a eu recours pour évaluer la qualité des données du recensement. Toutefois, la principale application de l'échantillonnage au recensement a été effectuée pendant le dénombrement sur le terrain où toutes les données, excepté les données de base, ont été recueillies auprès d'un échantillon de ménages. Le présent guide décrit cette dernière utilisation de l'échantillonnage et présente une évaluation de l'incidence de cette méthode sur la qualité des données du recensement.

Le chapitre II fait l'historique de l'application des méthodes d'échantillonnage aux recensements canadiens et décrit celles qui ont été utilisées au recensement de 1986. Le chapitre III décrit les méthodes de pondération des données-échantillon utilisées en vue d'obtenir des estimations à l'échelle de la population totale et explique les considérations d'ordre pratique et théorique qui régissent le choix de ces méthodes. On présente au chapitre IV les études effectuées dans le cadre du programme d'évaluation des méthodes d'échantillonnage et de pondération utilisées au recensement de 1986. Les chapitres V à VIII donnent les résultats de ces études.

II. APPLICATION DE L'ÉCHANTILLONNAGE LORS DE RECENSEMENTS CANADIENS

Dans un recensement de la population, l'échantillonnage consiste à recueillir et à traiter certaines caractéristiques à partir d'un échantillon aléatoire de logements et de personnes dénombrés lors du recensement intégral. On obtient ensuite pour l'ensemble de la population les totaux pour ces caractéristiques en pondérant les données-échantillon. Les caractéristiques recueillies pour tous les logements ou toutes les personnes faisant partie du recensement sont appelées "caractéristiques de base", tandis que les caractéristiques recueillies à partir d'un échantillon sont appelées "caractéristiques-échantillon".

A. Historique de l'échantillonnage au recensement canadien[1]

L'échantillonnage a été utilisé pour la première fois au cours du recensement de 1941. Un questionnaire sur le logement a été remis à tous les 10 logements de chaque sous-district de recensement. Les réponses aux 27 questions de ce questionnaire ont été intégrées à celles du questionnaire sur la population et les ménages correspondant à ces logements, ce qui a permis de faire le recoupement des caractéristiques-échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage au cours de l'étape du dépouillement pour obtenir des estimations provisoires du revenu des salariés, de la répartition de la population active selon l'âge et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du recensement de 1961, on a prélevé un échantillon de 20% des ménages privés et on a demandé aux membres de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. La méthode de l'échantillonnage n'a pas été utilisée au cours des recensements de moindre envergure de 1956 et 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de recensement en 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par représentant. Ce changement découle des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964, Hansen et associés, 1959) selon lesquels la présence d'un représentant avait un effet important sur la variance[2] des chiffres du recensement. On prévoyait que l'autodénombrement permettrait

---

1 Des renseignements détaillés sur les recensements particuliers sont contenus dans le rapport administratif, la revue générale, le guide sommaire ou le recensement en bref des recensements en question. Ces rapports sont mentionnés à la fin de ce guide.

2 La "variance" d'une estimation est une mesure de sa précision. La variance est expliquée plus en détail au chapitre VIII.

de réduire la variance des chiffres du recensement attribuable à la présence du représentant et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire à son aise; par conséquent, on pouvait s'attendre que les réponses seraient plus exactes que par le passé.

Un autre changement survenu dans le recensement de 1971 a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements précédents. Une étude des coûts, du fardeau du répondant et des délais de production par rapport à la qualité des données dans le cadre d'un recensement par échantillonnage et par autodénombrement a présidé à la décision de recueillir toutes les données, sauf celles portant sur quelques caractéristiques de base, auprès d'un échantillon d'un tiers de la population au recensement de 1971. Sauf dans les régions les plus isolées du Canada, le tiers des ménages privés a reçu le questionnaire complet comportant toutes les questions du recensement, tandis que le reste des ménages a reçu le questionnaire abrégé qui ne contenait que les questions de base soit le nom, le lien de parenté avec le chef de ménage, le sexe, la date de naissance, l'état matrimonial, la langue maternelle, le type de logement, le mode d'occupation, le nombre de pièces, l'alimentation en eau, les toilettes ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions prédésignées et dans des logements collectifs[3] ont reçu le questionnaire complet. Le rapport statistique "Sampling in the Census" (Bureau fédéral de la statistique (1968)) donne une description détaillée des projets d'application de cette méthode au recensement de 1971.

Le contenu du recensement de 1976 était beaucoup moindre que celui de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté à recueillir, par exemple, les questions sur le revenu, ni celles dont les frais de codage sont les plus élevés, par exemple, celles sur la profession, l'activité économique, et le lieu de travail. Par conséquent, les avantages de l'échantillonnage par rapport aux coûts et à l'allègement du fardeau du répondant étaient moins évidents qu'au recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles reliées à divers taux d'échantillonnage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral, il a été décidé d'appliquer en 1976 la même méthode d'échantillonnage qu'en 1971.

Au recensement de 1981, on a utilisé la plupart des mêmes méthodes qu'en 1971 et 1976, sauf que le taux d'échantillonnage est passé du tiers au cinquième des ménages. Les études réalisées à cette époque ont révélé que la diminution consécutive de la qualité des données (mesurée au moyen de la variance) serait acceptable et qu'elle ne serait pas suffisamment importante pour annuler les avantages découlant de la réduction des coûts et du fardeau du répondant, ainsi que la diminution des délais de production (voir Royce,

---

[3] Un logement collectif est une institution ou un établissement commercial ou communautaire, par exemple les hôtels, les hôpitaux, les résidences de personnel, et les camps de chantier.

1983). Douze questions ont été posées à l'ensemble de la population et 34 questions de plus ont été posées aux membres de l'échantillon.

Le recensement de 1986 a été le premier recensement majeur réalisé au milieu d'une décennie. Il avait été décidé que seul un recensement majeur pourrait répondre aux besoins croissants en matière de données locales sur le marché du travail, besoins rendus plus pressants par l'importante récession (1981-1982) survenue depuis le recensement précédent. Toutefois, afin de maintenir les coûts d'élaboration au niveau le plus bas possible, on a décidé de faire très peu de changements. Par conséquent, les questions ainsi que les méthodes de collecte et de dépouillement des données du recensement de 1981 ont été reprises, sauf lorsque nous avions des raisons importantes de ne pas procéder ainsi. Des questions portant sur huit sujets posées au recensement de 1981 n'ont pas été reprises en 1986, tandis que trois nouvelles questions ont été ajoutées.

B. Plan d'échantillonnage utilisé lors du Recensement de 1986

Le jour du recensement en 1986, une multitude de renseignements ont été recueillis auprès de chaque canadien. La majeure partie de ces informations ont été obtenues à partir d'un échantillon. Dans chaque secteur d'autodénombrement, 1 logement privé occupé sur 5 a été choisi pour répondre au questionnaire complet (questionnaire 2B) contenant toutes les questions du recensement, à savoir neuf questions de base sur l'âge, le sexe, l'état matrimonial, la langue maternelle, le lien de parenté avec la personne repère du ménage (personne 1), le genre de logement, le mode d'occupation; quatre questions de plus sur le logement et 19 questions de nature socio-économique. Le reste des ménages privés a reçu un questionnaire abrégé (questionnaire 2A) contenant seulement les neuf questions de base.

Tous les logements se trouvant dans les secteurs dénombrés par un recenseur (généralement les régions éloignées ou les réserves indiennes) ont reçu le questionnaire 2B. Tous les logements collectifs ont aussi reçu le questionnaire 2B. Toutefois, les personnes décrites ci-après habitant dans des logements collectifs n'ont pas eu à répondre au questionnaire-échantillon:

(a) les détenus dans les établissements de correction, les établissements pénitentiaires ou les prisons;

(b) les patients dans les hôpitaux généraux, les centres de soins spéciaux, les établissements pour personnes âgées et malades chroniques, ou les hôpitaux psychiatriques;

(c) les enfants dans les orphelinats et foyers pour enfants ou les établissements pour jeunes délinquants.

Les canadiens postés à l'extérieur du Canada (généralement le personnel des ambassades ou des forces armées) ont reçu un questionnaire 2C qui contenait les mêmes questions que le questionnaire 2B, à l'exclusion de celles portant sur le logement. Toutefois, ce questionnaire comportait des questions sur le lieu habituel de résidence de la personne au Canada. Les renseignements relatifs aux logements privés inoccupés ont été consignés sur le questionnaire 2A.

D'après les procédures de livraison, le représentant du recensement (RR) devait planifier un itinéraire couvrant tous les logements de son secteur de dénombrement (SD) puis se rendre à chaque logement et y laisser un questionnaire. Le choix de l'échantillon, c'est-à-dire la décision quant au genre de questionnaire à laisser dans chaque logement occupé, a été fait à partir du registre des visites (RV) dans lequel le RR inscrivait tous les logements de son secteur. Chaque cinquième ligne du registre était ombrée, indiquant qu'un questionnaire 2B devait être laissé. Un point de départ aléatoire a été déterminé à partir de la première ligne du RV, le RR sautait zéro, une, deux, trois ou quatre lignes selon que le cinquième, quatrième, troisième, deuxième ou premier logement du SD devait être le premier à recevoir le questionnaire complet. Par la suite, chaque logement inscrit sur une ligne ombrée recevait automatiquement le questionnaire complet. Ces procédures étaient expliquées dans le manuel du RR et une attention particulière y était portée pendant la formation du RR afin de minimiser tout risque d'erreur dans la sélection de l'échantillon.

Dans la terminologie de l'échantillonnage, l'échantillon est défini comme un échantillon stratifié avec sélection systématique de logements privés occupés, choisi en fonction d'un taux d'échantillonnage constant de 1 sur 5 dans toutes les strates (SD). Dans le cas de l'échantillon de personnes, il peut être défini comme étant un échantillon stratifié avec sélection systématique de grappes, un logement constituant une grappe. Stuart (1976) et Cochran (1977) donnent une description détaillée des concepts et de la terminologie se rapportant à l'échantillonnage.

C. <u>Traitement des données de l'échantillon du Recensement</u>

Après que le RR eut reçu le questionnaire rempli (2A ou 2B) de chaque logement de son secteur et que ce travail eut été approuvé, les questionnaires ont été envoyés à l'un des sept sites de dépouillement régionaux aux fins du traitement manuel. Les données complètes pour chaque SD ont été saisies et stockées sur rubans magnétiques. Les questionnaires et les rubans magnétiques ont ensuite été envoyés au bureau central, à Ottawa, où ont été effectuées par ordinateur des vérifications afin de déceler dans les données des incohérences qui ne pourraient être corrigées que par un examen manuel du questionnaire. Une fois terminées toutes les mises à jour des données pour un SD, les données étaient restructurées et transférées à la section Contrôle et Imputation.

Les données ont été chargées dans 10 bases de données de Contrôle et Imputation, organisées par données 2A (100%) et 2B (20%) pour chacune des 5 régions. Les bases de données 2A contenaient les caractéristiques démographiques de base pour l'ensemble de la population et les bases de données 2B contenaient les données pour les questions-échantillon. Les données ont été soumises à une série de modules de traitement spécialisés, où tous les problèmes liés à des données incohérentes, incompatibles, et manquantes ont été résolus. Les bases de données 2A ont été traitées en premier et ont donné lieu à la création de la base de données d'extraction 2A définitive.

Une fois la dernière main mise aux données intégrales, les réponses du questionnaire-échantillon ont été traitées. Les enregistrements 2B pour lesquels il y a eu non réponse ont été enlevés des bases de données 2B. Une base de données d'extraction 2B définitive a été créée. Cette base de données contenait à la fois les données intégrales et les données-échantillon uniquement pour les ménages et les personnes de l'échantillon. Les coefficients

de pondération calculés à l'aide des données intégrales (tels que décrits au chapitre III) ont été chargés dans cette base de donnée.

III. ESTIMATIONS BASÉES SUR L'ÉCHANTILLON DU RECENSEMENT

Tout échantillonnage nécessite l'application d'une méthode d'estimation pour pondérer les données-échantillon en vue d'obtenir des estimations à l'échelle de la population totale. La méthode d'estimation est généralement choisie en fonction d'exigences d'ordre pratique et théorique. Sur le plan pratique, cette méthode doit être compatible avec le système de dépouillement auquel elle s'intègre, tandis que sur le plan théorique, elle doit minimiser l'erreur des estimations due à l'échantillonnage. Les deux sections qui suivent décrivent les considérations pratiques et théoriques qui régissent le choix des méthodes d'estimation à partir de l'échantillon du recensement.

A. Considérations d'ordre pratique

Une méthode d'estimation peut être décrite mathématiquement à l'aide d'une formule algébrique qui indique comment la valeur de l'estimateur de la population est calculée en fonction des valeurs observées dans l'échantillon. Quand il s'agit de petites enquêtes qui ne portent que sur une ou deux caractéristiques, ou, lorsque la formule d'estimation est très simple, il est possible de calculer les estimations-échantillon en appliquant ladite formule aux données-échantillon pour chaque estimation requise. Toutefois, dans le cas d'une enquête ou d'un recensement portant sur une gamme étendue de caractéristiques, ou lorsque la formule d'estimation est assez complexe, il n'est pas possible d'appliquer une formule séparément pour chaque estimation désiré. Dans le cas du recensement, par exemple, chaque cellule de chaque totalisation, fondée sur des données-échantillon recueillies à chaque niveau géographique, représente une estimation-échantillon qui nécessite une application distincte de la formule d'estimation. Par ailleurs, le fait de calculer séparément chaque estimation ne mènerait pas nécessairement à la cohérence des diverses estimations effectuées à partir d'un même échantillon.

Par conséquent, la méthode utilisée pour le recensement et pour de nombreuses enquêtes-échantillon consiste à décomposer le processus d'estimation en deux étapes: (a) le calcul des coefficients de pondération, c'est-à-dire la pondération elle-même et (b) la sommation des coefficients de pondération afin de produire des chiffres estimés de la population.

Les seules difficultés mathématiques, s'il y en a, se rapportent à l'étape (a) qui n'est effectuée qu'une fois, tandis que l'étape (b) se limite à un simple processus de sommation des coefficients de pondération au moment où une totalisation est extraite. De plus, comme le coefficient de pondération relié à une unité-échantillon est le même quelle que soit la totalisation effectuée, la cohérence entre les diverses estimations basées sur les données-échantillon est assurée.

B. Considérations d'ordre théorique

Selon la théorie de l'échantillonnage, il est possible de déterminer, à partir d'un plan d'échantillonnage particulier et d'une procédure d'estimation donnée, les chances qu'un certain intervalle contienne la valeur (inconnue) que l'on veut estimer pour la population entière. Le principal objet de l'estimation est de réduire au maximum l'étendue de ces intervalles de façon à ce que les hypothèses sur les valeurs inconnues de la population soient aussi précises que possible. La mesure courante de précision utilisée pour comparer

les procédures d'estimation est appelée l'erreur-type. Pourvu que certaines conditions relativement peu rigoureuses soient respectées, un intervalle de plus ou moins deux erreurs-types par rapport à la valeur estimée renfermera la valeur de la population pour environ 95% de tous les échantillons possibles.

En plus de minimiser l'erreur-type, le deuxième objectif du choix d'une méthode d'estimation de l'échantillon du recensement est d'assurer autant que possible que les estimations-échantillon des caractéristiques de base soient cohérentes avec les valeurs connues de la population correspondante. Heureusement, ces deux objectifs sont généralement complémentaires, c'est-à-dire que l'erreur d'échantillonnage est habituellement réduite lorsqu'on s'assure que certaines données de base sont compatibles avec les chiffres correspondants de la population. Cependant, bien que cela soit vrai en général, le fait de forcer les estimations-échantillon des caractéristiques de base à être cohérentes avec les chiffres correspondants de la population de très petites cellules peut avoir de graves conséquences sur l'erreur d'échantillonnage des estimations des caractéristiques-échantillon proprement dites.

Lorsqu'on n'a aucune information sur la population échantillonnée autre que celles qui ont été recueillies pour les unités d'échantillonnage, la méthode d'estimation se limite à pondérer les unités d'échantillonnage inversement à leur probabilité de sélection, c'est-à-dire si toutes les unités d'une province avaient une chance sur 5 d'être choisies, toutes les unités de cette province recevraient un coefficient de pondération de 5. En pratique, on a généralement quelque connaissance de la population observée, par exemple, sa taille globale et, éventuellement, subdiviser en fonction d'une variable donnée, supposons, la province. Ces données peuvent être utilisées pour améliorer la formule d'estimation de façon à produire des estimations qui se rapprochent davantage de la valeur inconnue de la population. Dans le cas de l'échantillon du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée, soit les données intégrales, à tous les niveaux géographiques. D'une part, on peut se servir de cette information pour améliorer les estimations faites à partir de l'échantillon du recensement mais d'autre part, cette multitude de données peut également occasionner des ennuis compte tenu du fait qu'il est impossible d'obtenir des estimations-échantillon des caractéristiques de base qui soient cohérentes avec toutes les données sur la population à tous les niveaux géographiques. Les écarts entre les estimations-échantillon et les valeurs de la population apparaissent quand on produit un recoupement d'une variable intégrale et d'une variable-échantillon. La totalisation doit être fondée sur les données-échantillon; de ce fait, les totaux marginaux des variables intégrales, sont des estimations-échantillon, que l'on peut comparer avec les chiffres correspondants de la population d'une autre totalisation basée sur les données intégrales. Il se peut qu'ils ne concordent pas exactement.

C. Élaboration d'une méthode d'estimation pour l'échantillon du recensement

Étant donné qu'il faut attribuer un coefficient de pondération à chaque unité (personne, famille ou ménage) dans l'échantillon, le plus simple serait de choisir le coefficient 5, puisqu'il s'agit d'un échantillon de 1 sur 5. Une telle méthode serait simple et sans biais[4], et si l'on n'avait aucune autre donnée que celles de l'échantillon, cela pourrait être la meilleure méthode. Or, bien qu'on sache que l'échantillon contiendra presque exactement le cinquième de tous les ménages (à l'exclusion des ménages collectifs et de ceux qui se trouvent dans les secteurs dénombrés par un représentant), on ne peut pas être certain qu'il contiendra exactement un cinquième de toutes les personnes, ou un cinquième de chaque genre de ménage, ou un cinquième de toutes les femmes de 25 à 34 ans, etc. Par conséquent, une telle méthode n'assurerait pas la cohérence, même s'il s'agit des sous-groupes les plus importants de la population. Dans ce dernier cas, les fractions seraient très proches d'un cinquième, mais elles pourraient varier de façon sensible dans le cas des petits sous-groupes. Une autre méthode simple serait de prendre certains sous-groupes fondamentaux, par exemple, les groupes d'âge-sexe par province, puis, pour chacun, compter le nombre d'éléments de la population appartenant au sous-groupe (N) et le nombre d'unités de l'échantillon appartenant au sous-groupe (n), et d'attribuer à chaque unité d'échantillonnage un coefficient de pondération égal à N/n.

Par exemple, si l'on recensait 5 000 hommes de 20 à 24 ans dans l'Île-du-Prince-Édouard, et si 1 020 de ces hommes faisaient partie des ménages inclus dans l'échantillon, un coefficient de pondération de 5 000/1 020 = 4.90 serait attribué à chaque homme de 20 à 24 ans dans l'échantillon de l'Île-du-Prince-Édouard. De cette façon, chaque fois que le groupe d'âge quinquennal serait recoupé selon l'âge et le sexe avec une caractéristique-échantillon de l'Île-du-Prince-Édouard, le total marginal pour le groupe d'âge-sexe d'hommes de 20 à 24 ans concorderait avec le total de la population, soit 5 000. À noter que, dans ce cas, un coefficient de pondération de 5 donnerait une estimation-échantillon de 5 100 (1 020 x 5).

Cette méthode est connue sous le nom "d'estimation par la méthode du quotient". On peut prouver que, dans de nombreux cas, son utilisation mène à une forte réduction de l'erreur-type. Cette méthode permet d'assurer la cohérence entre les estimations-échantillon, les totaux de population des sous-groupes choisis et des combinaisons de ces sous-groupes. Cependant, elle ne permet pas d'assurer la cohérence pour les plus petits sous-groupes, par exemple, par année d'âge dans les comtés, ou pour d'autres groupes définis selon d'autres caractéristiques de base, par exemple, l'état matrimonial ou la langue maternelle. On pourrait envisager d'étendre cette méthode à de plus petits sous-groupes qui seraient définis comme les cellules dans un recoupement de l'ensemble des caractéristiques de base pertinentes. Un problème se pose alors, car à mesure que le sous-groupe devient plus petit, cette méthode devient instable, c'est-à-dire que les erreurs-types des estimations ainsi produites augmentent. À la limite, cette méthode devient impossible à utiliser lorsqu'un sous-groupe particulier ne comprend plus aucune unité-échantillon. Le défi consiste donc

---

4 "Sans biais" veut dire que la moyenne des estimations de l'ensemble de tous les échantillons possibles serait égale à la valeur de la population réelle.

à profiter des avantages d'une estimation par la méthode du quotient sans subir l'instabilité reliée à l'utilisation de petits sous-groupes. La solution adoptée consiste à appliquer cette méthode de façon itérative à deux ensembles distincts et exhaustifs de sous-groupes. Cette méthode, appelée "méthode itérative par le quotient" (MIQ) et utilisée au recensement de 1986, est décrite ci-après.

D. Méthode itérative par le quotient

Dans cette méthode, on définit deux ensembles de sous-groupes au lieu d'un. Un des ensembles sert à former les lignes d'une "matrice de pondération" tandis que l'autre en forme les colonnes. Ainsi, dans le calcul des coefficients de pondération des personnes, les sous-groupes âge-sexe-état matrimonial forment les lignes de la matrice, tandis que les sous-groupes situation des particuliers dans la famille-langue maternelle en forment les colonnes.

Supposant qu'on ait la matrice appropriée, la MIQ se présente comme suit:

(a) On fait le classement recoupé des éléments de la population dans la matrice de façon à obtenir les totaux de la population de chaque ligne et de chaque colonne.

(b) On fait le classement recoupé des éléments de l'échantillon dans la même matrice pour obtenir les comptes d'échantillon pour chaque cellule et les totaux d'échantillon de chaque ligne et de chaque colonne.

(c) Si besoin est, on combine les lignes et les colonnes de la matrice en fonction de certaines limites de taille (voir ci-après).

(d) On affecte un coefficient de pondération initial de 5 à chaque élément de l'échantillon.

(e) Pour chaque colonne, on compare le total estimé à l'aide de ce coefficient de pondération initial au total connu de la population. On élimine tout écart en multipliant les coefficients de pondération initiaux par le rapport du total connu de la population au total estimé de la colonne. Les coefficients de pondération ainsi obtenus sont appelés coefficients de pondération de la première itération.

(f) Pour chaque ligne, on compare le total estimé à l'aide des coefficients de pondération de la première itération au total connu de la population de la ligne. On élimine tout écart en multipliant les coefficients de pondération de la première itération par le rapport du total connu de la population au total estimé de la ligne (la concordance exacte des colonnes s'en trouve détruite).

(g) On applique cette méthode d'élimination des écarts dans les estimations des colonnes et des lignes jusqu'à ce que les écarts qui restent soient négligeables (on dit alors qu'il y a eu "convergence") ou jusqu'à concurrence de 80 itérations.

On cesse l'application de la méthode avec les lignes de sorte que les totaux de celles-ci concordent exactement et que les totaux des colonnes concordent presque exactement. L'élément important de cette méthode est que les limites de taille visant à éviter l'instabilité des estimateurs s'appliquent seulement aux totaux des lignes et des colonnes et non aux

cellules particulières de la matrice, certaines d'entre elles pouvant d'ailleurs être vides. Pour obtenir plus de détails sur la MIQ, consulter Brackstone et Rao (1979). La MIQ a été élaborée à partir de la "méthode de l'ajustement proportionnel itératif", que Deming et Stephan (1940) ont été les premiers à présenter.

La MIQ comporte deux paramètres qui ont une importance cruciale par rapport à la cohérence entre les estimations-échantillon des caractéristiques de base et les totaux correspondants de la population. Le premier est le choix de la région géographique ou région de pondération (RP) à laquelle on applique cette méthode. Les étapes (a) à (g) décrites précédemment sont appliquées indépendamment à l'intérieur de chaque RP. Le deuxième est le choix des sous-groupes qui servent à former les lignes et les colonnes de la matrice de pondération.

La RP est cette région géographique pour laquelle une concordance presque exacte des totaux de personnes et de ménages ainsi que des sous-groupes formant les lignes et les colonnes de la matrice de pondération est assurée. Du point de vue de la cohérence dans le cas des petites régions, plus la RP est petite, plus la cohérence sera grande. Dans ce cas, le nombre de détails dans les lignes et les colonnes sera réduit à cause des limites de taille minimales de ces lignes et de ces colonnes. Lors du recensement de 1986, on a adopté le compromis suivant:

(a) une RP ne devrait contenir que 2 000 à 7 000 personnes (compte intégral);

(b) les limites des RP doivent respecter les limites des divisions de recensement (DR) et, dans la mesure du possible, celles des subdivisions de recensement (SDR), celles des secteurs de recensement (SR) ainsi que celles des circonscriptions électorales fédérales (CEF);

(c) Les RP devraient être composées de SD complets et devraient, de façon générale, être contiguës (c'est-à-dire ne présenter aucune "zone morte").

Il y a deux critères pour la sélection des sous-groupes d'une matrice de pondération. En premier lieu, la corrélation entre les variables qui définissent les sous-groupes et les caractéristiques-échantillon est importante pour minimiser l'erreur d'échantillonnage des estimations-échantillon des catégories. En second lieu, la nécessité d'assurer la cohérence de certains sous-groupes importants influence le choix des lignes et des colonnes. Ces deux critères sont souvent, mais pas nécessairement, complémentaires. La matrice ne peut pas être très détaillée à cause de la taille limite des lignes et des colonnes. En 1986, on a utilisé deux matrices distinctes, une pour les coefficients de pondération des personnes et des familles et l'autre, pour les coefficients de pondération des ménages. Ces deux matrices sont illustrées à l'annexe.

Pour chaque matrice, on a eu recours à une stratégie de regroupement afin de définir comment les lignes et les colonnes seraient combinées dans le cas des RP où une ligne ou une colonne présentent l'un des critères suivants :

(a) le total de la population de la ligne ou de la colonne était inférieur à 35;

(b) le chiffre de l'échantillon de la ligne ou de la colonne était égal à 0;

(c) le rapport du chiffre de population au chiffre de l'échantillon pour la ligne ou la colonne ne se situait pas dans l'intervalle de 3.0 à 19.9.

Les stratégies de regroupement ont été choisies de façon à préserver les sous-groupes chaque fois que cela était possible. Par exemple, au besoin, la colonne "logement loué - autres" dans la matrice des ménages a été combinée à la colonne "logement loué", et non avec une des colonnes "logement possédé", afin de conserver le sous-groupe "logement loué". Les regroupements ont été faits de façon indépendante dans chaque RP jusqu'à ce que toutes les contraintes relatives à la population des lignes et des colonnes ainsi qu'aux rapports aient été respectées. Dans les matrices présentées dans l'annexe, des lignes simples et doubles séparent les lignes et les colonnes en groupes. On a d'abord fait le regroupement dans les groupes séparés par des lignes simples, lorsqu'elles existaient. Puis, au besoin, on a fait les regroupements en franchissant les lignes simples, mais à l'intérieur des lignes doubles. On n'a fait que très rarement des regroupements en franchissant les lignes doubles. Cela ne s'est produit que lorsque toutes les lignes ou les colonnes situées entre deux lignes doubles avaient été regroupées en une ligne ou une colonne dont le chiffre de l'échantillon était toujours zéro mais dont le chiffre de population était supérieur à zéro.

La MIQ a produit des coefficients de pondération finals qui étaient les mêmes pour toutes les unités d'une même cellule de la matrice combinée, mais qui variaient d'une cellule à l'autre. Ces coefficients finals ont alors été ajoutés à l'enregistrement de chaque unité-échantillon dans la base de données. À chaque personne dans l'échantillon, on a attribué le coefficient de pondération calculé pour la cellule de la matrice des personnes et des familles à laquelle elle appartenait; à chaque ménage dans l'échantillon, on a attribué le coefficient de pondération associé à la cellule appropriée de la matrice des ménages; à chaque famille de recensement dans l'échantillon, on a associé le coefficient de pondération particulier de l'époux ou du parent seul de la famille. Le coefficient de pondération associé à l'époux de la personne repère de la famille économique ou, si l'époux n'était pas présent, le coefficient de pondération attribué à la personne repère de la famille économique a été attribué à la famille économique. Un coefficient de pondération égal à un a été automatiquement attribué à toutes les personnes, toutes les familles, et tous les ménages appartenant à des secteurs non échantillonnés.

En pratique, l'application de la MIQ s'est faite presque entièrement par ordinateur. Les RP ont été formées en utilisant un programme qui tenait compte de la population du SD, des coordonnées géographiques des centroïdes des SD et des secteurs géostatistiques (DR, SDR, etc.) auxquels les SD appartiennent. Ce programme a produit une liste énumérant les RP créées et permettait des modifications manuelles si nécessaire. On n'a fait des modifications manuelles que dans un nombre restreint de cas. Une fois les RP créées, on a utilisé des méthodes automatisées pour le recoupement des données, le regroupement des lignes et des colonnes, le calcul des coefficients de pondération et l'attribution de ces coefficients aux enregistrements de la base de données.

IV. PROGRAMME D'ÉVALUATION DE L'ÉCHANTILLONNAGE ET DE LA PONDÉRATION

Le programme d'évaluation de l'échantillonnage et de la pondération avait pour objet de déterminer l'effet de l'échantillonnage et de la pondération sur la qualité des données-échantillon du recensement. À cette fin, cinq études ont été réalisées afin de mesurer la qualité des données-échantillon et des estimations-échantillon du recensement ainsi que de fournir des renseignements utiles à la planification des futurs recensements. Les études ont porté sur les sujets suivants :

(a) examen du biais d'échantillonnage;

(b) évaluation de la formation des régions de pondération;

(c) évaluation des méthodes de pondération;

(d) examen de la cohérence des estimations-échantillon et des chiffres de population;

(e) étude visant à produire les estimations de la variance pour diverses caractéristiques-échantillon.

La suite du présent chapitre décrit brièvement chacune de ces études. Les chapitres V à VIII en présentent les résultats.

A. <u>Examen du biais d'échantillonnage</u>

Dans toute enquête, un biais entachant les réponses peut provenir de diverses sources. Le but de cette étude était de déterminer si les réponses aux questions de base du questionnaire 2B étaient biaisées de quelque façon que ce soit et de cerner, si possible, les causes de tout biais observé.

B. <u>Évaluation de la formation des régions de pondération</u>

Cette étude visait à mesurer le degré auquel les RP satisfaisaient aux critères de formation établis (voir le chapitre III, section D). Toutes les RP au Canada ont fait l'objet d'une analyse en vue de déterminer dans quelle mesure elles respectaient les contraintes de taille et les limites des diverses catégories de régions géographiques. On a examiné les raisons pour lesquelles des contraintes de taille n'étaient pas respectées.

C. <u>Évaluation des méthodes de pondération</u>

Le but de cette étude était d'évaluer l'efficacité de la MIQ. À cette fin, on a examiné le niveau de concordance entre les estimations-échantillon et les chiffres de population pour les lignes et les colonnes des matrices de recoupement de toutes les RP au Canada. Le nombre de regroupements de lignes et de colonnes, le degré de convergence de la MIQ, et la variabilité des taux d'échantillonnage et les tailles de population parmi les lignes et les colonnes ont aussi été examinés afin d'expliquer les incohérences observées.

D. Examen de la cohérence des estimations-échantillon et des chiffres de population

Dans cette étude, on a examiné le niveau de concordance, ou la cohérence, entre les estimations-échantillon et les chiffres de population pour un large éventail de caractéristiques de base, et non seulement pour celles utilisées pour définir les lignes et les colonnes des matrices de recoupement. On a étudié la concordance pour diverses catégories de régions géographiques autres que les RP, dont les limites ne respectaient pas toujours celles des RP. De plus, on a examiné la cohérence des cellules individuelles des matrices de recoupement. Une étude distincte a été effectuée en vue d'évaluer la cohérence dans le cas de la caractéristique "langue maternelle".

E. Étude de la variance d'échantillonnage

La variance d'une estimation est une sorte de mesure de sa précision. Il est relativement peu coûteux de calculer des estimations de la variance pour des estimations lorsqu'on utilise des coefficients de pondération égaux à 5 et que l'on suppose un échantillonnage aléatoire simple. Toutefois, il est beaucoup plus onéreux d'estimer la variance, dans le cas d'estimations établies par la MIQ, en prenant en considération le plan d'échantillonnage utilisé. Le but de cette étude était d'élaborer une méthode peu coûteuse de production de telles estimations de la variance. Pour ce faire, on a calculé des "facteurs d'ajustement". Un facteur d'ajustement est le rapport entre l'estimation de l'erreur-type (la racine carré de la variance) associée à la MIQ et l'estimation de l'erreur-type simple. Une estimation de l'erreur type associée à la MIQ pour n'importe quelle caractéristique dans toute région géographique peut alors être obtenue en multipliant l'estimation de l'erreur-type simple par le facteur d'ajustement approprié.

V. BIAIS D'ÉCHANTILLONNAGE

Les estimations établies à partir d'une enquête-échantillon sont sujettes à des erreurs d'échantillonnage. Un type d'erreur d'échantillonnage est causé par la variabilité des caractéristiques de la population. Cette variabilité signifie que des échantillons différents produiront des estimations différentes, aucune d'elles n'égalant nécessairement la vraie valeur de la population. Toutefois, les estimations égaleront la vraie valeur de la population en moyenne pourvu que l'échantillon ne comporte pas de biais donnant lieu à une tendance à la surestimation ou à la sous-estimation. Malheureusement, il est souvent difficile d'éliminer le biais complètement. Dans le cas du recensement de la population, les réponses peuvent être entachées d'un biais causé, entre autres, par les erreurs de couverture, les erreurs dues à la non-réponse, la déviation des réponses (par ex., des recensés qui répondent différemment sur la formule 2B que sur la formule 2A), les erreurs commises par les recenseurs (par ex., des recenseurs ne choisissant pas l'échantillon selon les exigences établies), les erreurs de traitement, et ainsi de suite.

L'examen du biais d'échantillonnage a été effectué pour déterminer si les réponses aux questions de base de la formule 2B comportaient un déviation. Les estimations-échantillon d'un grand nombre de caractéristiques de base ont été comparées aux chiffres de population pour les 260 divisions de recensement (DR) échantillonnées au Canada. Ces estimations-échantillon ont été obtenues en multipliant les chiffres de l'échantillon au niveau du SD par le poids simple qui est égal à la fraction de sondage inverse du SD (environ 5) et en faisant ensuite la somme au niveau de la DR[5]. Des représentations graphiques des différences entre les estimations-échantillon et les chiffres de population pour chaque DR ont été produites séparément pour chaque caractéristique afin de déterminer l'existence d'aspects indiquant une tendance précise à produire des estimations trop faibles (biais par défaut) ou trop élevées (biais par excès). En outre, des tests ont été effectués pour déterminer si les différences entre les estimations-échantillon et les chiffres de population étaient statistiquement significatives.

La nature des différences qui ressortent des représentations graphiques indique que la plupart des caractéristiques de l'échantillon présentent effectivement un certain biais. De plus, la différence moyenne entre les estimations-échantillon et les chiffres de population calculée sur l'ensemble des DR a été jugée statistiquement significative (au niveau de 5%[6]) pour la plupart des caractéristiques (c'est-à-dire que les différences ne peuvent pas être expliquées par la variabilité de l'échantillonnage). Le tableau 1 indique les différences (chiffres absolus et pourcentages) entre les estimations-échantillon et les chiffres de population à l'échelle nationale (moyenne établie sur l'ensemble des DR) pour l'ensemble des caractéristiques étudiées. Dans la plupart des cas, le biais était inférieur à 1%. Le tableau montre aussi le pourcentage de DR dans lesquelles chaque caractéristique était sur-représentée. Un pourcentage inférieur à 50

---

[5] Ces estimations simples ont été utilisées au lieu des estimations obtenues par la méthode itérative du quotient étant donné que la MIQ permet de réduire le biais d'échantillonnage en forçant les estimations à égaler les chiffres de population.

[6] Cela signifie que la possibilité de relever de si grandes différences en l'absence d'un biais n'était qu'au plus 5%.

Tableau 1. Estimations-échantillon (pondération simple) moins la population totale à l'échelle nationale (SD échantillonnés seulement) et le pourcentage de DR où il y a surreprésentation des caractéristiques

Caractéristique	Estimations-échantillon moins la population totale			Pourcentage de DR où il y a surreprésentation
Caractéristiques - Personnes				
Hommes	1 557	*	(+0.01%)	54
Femmes	22 017	*	(+0.18%)	73
Population totale	23 574	*	(+0.10%)	71
Âge, 0-5	9 185	*	(+0.44%)	69
Âge, 6-14	17 787	*	(+0.57%)	75
Âge, 15-24	- 8 385	*	(- 0.21%)	43
Âge, 25-34	- 1 332	*	(- 0.03%)	56
Âge, 35-44	5 661	*	(+0.16%)	60
Âge, 45-54	2 844	*	(+0.11%)	57
Âge, 55-64	2 360		(+0.10%)	49
Âge, 65 ans et plus	- 4 546	*	(- 0.19%)	33
Célibataires	- 462	*	(- 0.00%)	55
Personnes mariées	37 195	*	(+0.32%)	83
Veufs et veuves	- 5 405	*	(- 0.49%)	29
Personnes divorcées	- 3 937	*	(- 0.59%)	41
Personnes séparées	- 3 817	*	(- 0.77%)	39
Caractéristiques - Familles de recensement				
Nombre total de familles	20 056	*	(+0.30%)	85
Familles époux-épouse	20 250	*	(+0.35%)	85
Familles monoparentales	- 194		(- 0.02%)	45
Enfants dans les familles	30 418	*	(+0.36%)	76
Personnes dans les familles	70 724	*	(+0.34%)	84
Personnes hors famille	- 47 150	*	(- 1.35%)	9
Caractéristiques - Ménages et logements				
Propriétaires de logements	5 995	*	(+0.11%)	60
Logements loués	- 5 995	*	(- 0.18%)	40
Maisons individuelles	6 792	*	(+0.13%)	68
App. dans un immeuble < que 5 étages	- 5 315	*	(- 0.31%)	31
App. dans un immeuble ≥ que 5 étages	- 209		(- 0.03%)	43
Logements mobiles	0		(+0.08%)	51
Autres logements	- 1 358	*	(- 0.12%)	42
Ménages d'une personne	- 13 248	*	(- 0.70%)	22
Ménages de 2 personnes	4 082		(+0.15%)	49
Ménages de 3 personnes	2 413	*	(+0.15%)	57
Ménages de 4 ou 5 personnes	8 459	*	(+0.36%)	72
Ménages de 6 personnes et plus	- 1 706		(- 0.52%)	47
Ménages non familiaux	- 21 249	*	(- 0.92%)	14
Ménages unifamiliaux	22 349	*	(+0.35%)	87
Ménages multifamiliaux	- 1 100	*	(- 1.17%)	39
Soutiens du ménage < 25 ans	- 3 739	*	(- 0.71%)	40
Soutiens du ménage 25-34 ans	1 673	*	(+0.08%)	59
Soutiens du ménage 35-44 ans	3 385	*	(+0.17%)	59
Soutiens du ménage 45-64 ans	2 302		(+0.09%)	57
Soutiens du ménage > 64 ans	- 3 621	*	(- 0.23%)	30
Soutiens du ménage sexe masculin	1 025		(+0.02%)	52
Soutiens du ménage sexe féminin	- 1 025		(- 0.04%)	48

* Ces différences sont statistiquement significatives au niveau 5%.

signifie que la caractéristique a été sous-représentée dans la majorité des DR. Les conclusions faites à partir de l'analyse des différences sont données dans les paragraphes qui suivent.

La taille des ménages de l'échantillon 2B était plus importante en moyenne que celle qui a été établie à partir des données sur l'ensemble de la population. Les groupes suivants avaient définitivement tendance à être sur-représentés dans l'échantillon: les femmes, les groupes d'âges 0-5 ans, 6-14 ans, 35-44 ans, et 45-54 ans, ainsi que les membres des familles de recensement, particulièrement les personnes mariées et les enfants des familles de recensement.

Les groupes de personnes ci-après étaient sous-représentés dans l'échantillon: les personnes de 15 à 24 ans et de plus de 64 ans, les veufs(veuves), les personnes séparées, et les personnes divorcées, et les personnes hors famille de recensement. La sous-représentation des personnes présentant ces caractéristiques est particulièrement significative compte tenu du fait qu'en moyenne, il y avait plus de gens dans les logements échantillonnés que dans les logements non échantillonnés.

En ce qui a trait aux caractéristiques des ménages, les logements possédés et les logements individuels non attenants avaient tendance à être sur-représentés dans l'échantillon tandis que l'inverse se produisait dans le cas des logements loués et des appartements situés dans des immeubles d'appartements de moins de 5 étages. Les ménages avec une famille de recensement tendaient à être sur-représentées tandis que les ménages hors famille de recensement et, dans une moindre mesure, les ménages formés de plusieurs familles de recensement étaient sous-représentés. De même, on a constaté une sur-représentation des ménages de quatre ou cinq personnes et une sous-représentation de ceux d'une personne. Les soutiens de ménage de 25-34 ans et de 35-44 ans étaient sur-représentés tandis que ceux de moins de 25 ans et de plus de 64 ans étaient sous-représentés.

Comme il a été mentionné précédemment, les différences observées entre les estimations-échantillon (basées sur une pondération simple) et les chiffres de population peuvent être expliquées de diverses façons. Il est possible qu'elles soient attribuables au fait qu'il y a eu 67 884 ménages n'ayant fourni aucune réponse (0.8% de l'ensemble) lors du recensement de 1986. Il s'agissait soit de ménages ayant complètement refusé de répondre aux questions ou pour lesquels le recenseur avait été incapable d'obtenir des renseignements (habituellement parce que les membres du ménage étaient absents au moment où le recensement a été mené ou avaient déménagé, le jour même du recensement ou après cette date, sans répondre au questionnaire). Le pourcentage de ménages échantillonnés constituant des cas de non-réponse était plus du double de celui des ménages non échantillonnés. Il est possible que les ménages non répondants aient présenté de façon générale des caractéristiques différentes de celles des ménages répondants (par ex., ménages de plus petite taille). Le cas échéant, ces différences auraient pu influer sur les données-échantillon de façon disproportionnée. Un biais attribuable à la non-réponse aurait été introduit aussi bien dans les données-échantillon que dans les données intégrales, et le fait que le biais aurait été plus important pour les ménages de l'échantillon aurait entraîné des divergences entre les estimations-échantillon et les chiffres de population.

Durant le traitement des données, les ménages échantillonnés non répondants (pour qui aucune question a été répondue) ont été supprimés de l'échantillon (de sorte que ces ménages sont

devenus des ménages non échantillonnés) et seules les réponses aux questions de base ont été imputées. Par conséquent, si dans le cadre du processus d'imputation certains types de ménages avaient tendance à se faire attribuer des réponses plus souvent que d'autres, cela aurait aussi donné lieu à des écarts entre les estimations-échantillon et les chiffres de population étant donné que seuls les ménages non échantillonnés auraient été touchés. Les données intégrales auraient alors été faussées, mais non les données-échantillon. Lorsque les ménages non répondants ont été supprimés de l'étude, on a constaté une certaine réduction du biais observé pour la plupart des caractéristiques. Toutefois, le biais restant était quand même statistiquement significatif au niveau de 5% pour 33 des 44 caractéristiques observées (le biais était significatif pour 35 des caractéristiques avant l'exclusion des ménages non répondants). Par conséquent, l'incidence de l'imputation faite pour les ménages n'ayant fourni aucune réponse n'a pas été assez importante pour expliquer tout le biais observé.

Nous avons également examiné les effets de la non-réponse partielle et, dans les cas de non-réponse totale aux questions-échantillon, de la conversion des enregistrements 2B en enregistrements 2A. On considère qu'il y a non-réponse partielle quand un recensé refuse de répondre à certaines questions, pas à toutes. Dans des cas semblables, les réponses ont été imputées par le système. Un taux de non-réponse plus élevé sur les formules 2B que sur les formules 2A, ou vice versa, va entraîner des écarts entre les estimations-échantillon et les chiffres de population si certains types de ménages ou des personnes ont eu davantage tendance que d'autres à refuser de répondre à certaines questions ou si le système d'imputation a attribué certaines réponses plus souvent ou moins souvent qu'il se devait. La conversion des enregistrements 2B en enregistrements 2A va créer des écarts entre les estimations-échantillon et les chiffres de population si certains types de ménages ont été plus portés que d'autres à ne pas répondre aux questions-échantillon.

En éliminant des données les effets de ces deux facteurs, le biais observé a été réduit encore davantage en ce qui concerne la plupart des caractéristiques (cette réduction s'ajoute à celle obtenue par l'élimination de l'échantillon des ménages où il y a eu non-réponse totale). Toutefois, le biais restant est demeuré statistiquement significatif pour 32 des 44 caractéristiques étudiées. Par conséquent, bien que ces facteurs semblent avoir contribué au biais, une bonne partie de ce dernier n'a pu être expliqué.

Il est aussi possible qu'une autre source de biais provienne du fait que les personnes vivant dans les ménages échantillonnées ont été sous-dénombrées plus souvent que ceux des ménages non échantillonnés. On sait que les caractéristiques des personnes oubliées diffèrent de celles des personnes dénombrées. Un biais serait introduit dans les données intégrales aussi bien que dans les données-échantillon, mais comme le taux de sous-dénombrement des personnes dans les ménages échantillonnés est plus élevé, des divergences entre les estimations-échantillon et les chiffres de population seraient créées. Pour de plus amples renseignements sur la couverture dans le recensement de 1986, voir le Guide à l'intention des utilisateurs sur la qualité des données du Recensement de 1986: Couverture.

Pour de plus amples renseignements sur l'examen du biais d'échantillonnage, voir Rathwell (1990).

VI. ÉVALUATION DES MÉTHODES DE PONDÉRATION

A. Formation de régions de pondération (RP)

La première étape des méthodes de pondération a été la formation des RP. Le but visé était de créer des RP suffisamment grandes pour assurer l'application efficace de la MIQ (population d'au moins 2 000 personnes), mais assez petites pour respecter les limites du plus grand nombre possible de subdivisions de recensement (SDR), de secteurs de recensement (SR) et de circonscriptions électorales fédérales (CEF). De plus, les RP devaient être créées en tenant compte des limites de toutes les divisions de recensement (DR). Les SD échantillonnés ont été formés en 5 341 RP[7] comptant en moyenne 4 558 personnes (les personnes vivant dans des logements collectifs étant exclues). Sur les 5 341 RP, 5 229 (98%) avaient une population se situant entre 3 000 et 7 000. Des 112 RP qui restaient, 107 avaient une population variant entre 2 001 et 2 999, la population de 93% de ces RP allant de 2 501 à 2 999. Les cinq dernières RP avaient une population comprise entre 1 et 2 000. Celles-ci avaient été créées spécialement en vue de correspondre aux SD ayant des fractions de sondage extrêmes (près de 0 ou de 100%).

On a examiné dans quelle mesure les RP respectaient les limites de chacun des différents types d'unités géographiques, notamment les SR, les SDR dans les centres divisés en secteurs de recensement, les SDR dans les centres non divisés en secteurs de recensement et les CEF. Étant donné que les limites des DR ont toujours été respectées, il n'a pas été nécessaire de faire de vérification à ce niveau. Seule la partie échantillonnée des unités géographiques a fait l'objet d'un examen visant à déterminer si les limites avaient été respectées. Les unités géographiques qui ne contenaient pas de SD échantillonnés ont été exclues de l'étude.

Le tableau 2 montre dans quelle mesure les RP ont respecté les limites des SR, des SDR et des CEF. La première colonne montre le pourcentage d'unités géographiques ne contenant que des RP entières. La deuxième colonne indique le pourcentage d'unités géographiques qui étaient trop petites pour former des RP entières, mais étaient complètement englobées par une RP. La troisième colonne présente le pourcentage d'unités qui comprennent des parties de RP différentes.

[7] Six RP additionnelles ont été formées par le système automatisé, mais puisqu'elles ne contenaient aucun SD échantillonné, elles n'ont pas été utilisées par la MIQ.

Tableau 2. Mesure dans laquelle les régions de pondération respectent les diverses limites géographiques

Unités géographiques	Contenant seulement des RP entières	Englobées entièrement par une RP	Contenant des parties de RP différentes
Divisions de recensement	100%	0%	0%
Secteurs de recensement	56%	32%	12%
Subdivisions de recensement dans des centres divisés en secteurs de recensement	57%	32%	11%
Subdivisions de recensement dans des centres non divisés en secteurs de recensement	8%	85%	7%
Circonscriptions électorales fédérales	15%	0%	85%

Comme la MIQ est appliquée indépendamment à l'intérieur des RP, la concordance entre les estimations-échantillon et les chiffres de population est assurée seulement dans le cas des unités géographiques contenant uniquement des RP entières. La concordance n'est pas assurée pour les unités géographiques qui sont complètement englobées par une RP ou qui contiennent des parties de RP différentes.

Les RP respectaient les limites des SR et des SDR des centres divisés en secteurs de recensement presque aussi bien dans un cas que dans l'autre. En raison de la petite taille de la plupart des SDR dans les centres non divisés en secteurs de recensement, la majorité d'entre elles (85%) ont été englobées par une RP.

Seulement 15% des CEF ne contenaient que des RP entières. En outre, comme les CEF sont considérablement plus grandes que les RP, aucune d'elles se trouvait complètement comprise dans une partie d'une RP. La majorité des CEF (80%) contenaient entre 5 et 30 RP entières, et entre 1 et 10 RP partielles. Leurs limites n'ont donc pas été bien respectées au moment de la formation des RP. Étant donné que les limites des CEF n'ont aucun rapport avec celles des DR, des SR et des SDR, elles ne pourraient pas être mieux respectées sans courir le risque d'enfreindre davantage les limites de ces dernières.

Pour de plus amples renseignements sur cette étude, voir Daoust (1987).

B. Évaluation de la méthode itérative du quotient

Un des buts de la méthode de pondération consiste à réduire les écarts entre les chiffres de population définis par les lignes et les colonnes des matrices de pondération et les estimations-échantillon correspondantes. Ces écarts sont le résultat de la variabilité et du biais d'échantillonnage (voir le chapitre V). Certains écarts peuvent toutefois subsister même après l'application de la méthode de pondération. Une des principales causes de ces écarts est le regroupement des lignes et des colonnes avant la pondération. Cette mesure est prise afin de respecter les contraintes relatives à la taille qui permettent d'assurer que la méthode de pondération donne les résultats voulus (voir le chapitre III, section D). Les écarts correspondent à la différence entre les estimations-échantillon et les chiffres de population, laquelle est exprimée en tant que pourcentage des chiffres de population, c'est-à-dire:

$$\text{écart} = \frac{\text{estimations-échantillon - chiffres de population}}{\text{chiffres de population}} \times 100$$

L'écart peut devenir important s'il y a combinaison d'une ligne de faible valeur et d'une ligne de valeur importante, ces deux lignes comportant des fractions de sondage très différentes.[8] Par exemple, supposons qu'une ligne qui a une population de 30 et un échantillon de 5 (c'est-à-dire un taux d'échantillonnage de 16.7%) est combinée à une ligne qui a une population de 90 et un échantillon de 19 (soit un taux d'échantillonnage de 21.1%). Le résultat est une ligne combinée qui a une population de 120 et un échantillon de 24. Le coefficient de pondération de cette ligne pourrait être définie comme étant 120/24 = 5, qui est le rapport des chiffres de population à l'estimation-échantillon (il s'agit, bien entendu, d'une simplification des résultats de l'application de la MIQ). Selon ce coefficient de pondération, l'estimation de la plus petite ligne serait 5 x 5 = 25, ce qui produirait un écart de (25 - 30)/30 x 100 = -16.7%. L'application du coefficient de pondération à la ligne ayant une valeur plus importante donnerait lieu à une estimation de 19 x 5 = 95. L'écart correspondrait à (95 - 90)/90 x 100 = 5.6%. Il est donc considérable pour la ligne ayant la valeur moins élevée et plus faible pour celle ayant la valeur plus forte.

Les écarts ont été calculés pour chaque ligne et chaque colonne de la matrice des personnes et de celle des ménages pour chacune des principales régions du pays (l'Est, le Québec, l'Ontario et l'Ouest, y compris les Territoires). Les écarts relevés pour toutes les lignes de la matrice des ménages étaient inférieurs à ± 4% pour toutes les régions du Canada. Par exemple, la figure 1 (page 23) montre les écarts observés pour le Canada et l'Est du pays. L'axe vertical représente l'écart exprimé en pourcentage et établi selon la formule précitée. Les valeurs de l'axe horizontal correspondent à celles des lignes de la matrice des ménages, comme on peut le voir dans l'annexe. Les lignes verticales montrent où se trouvent les lignes doubles sur la matrice, comme l'indique l'annexe; on regroupe rarement des colonnes se trouvant d'un côté et de l'autre de ces lignes (voir le chapitre III,

---

[8] Les différences entre les fractions de sondage sont causées par des facteurs tels que la variabilité d'échantillonnage, les corrections faites pour tenir compte des cas de non-réponse, le biais d'échantillonnage ou la déviation des réponses.

section D). Dans de nombreux cas, les écarts importants ont été causés par la combinaison de lignes de faible valeur et de lignes de valeur plus importante ayant des fractions de sondage très différentes. Par exemple, on peut voir d'après la figure 1 que la ligne 22 présente un écart d'environ - 2% pour l'Est. La ligne 22 représente les ménages non familiaux d'une personne dont le soutien du ménage est un homme de 65 ans ou plus. La figure 2 montre le pourcentage de la population (dans les SD échantillonnés seulement) s'appliquant à chaque ligne. On peut constater qu'il y a seulement environ 2% des ménages de la région de l'Est dans cette ligne. La figure 3 donne le pourcentage de RP pour lesquelles les lignes ont été combinées. La ligne 22 a été combinée à une autre ligne dans presque 80% des cas pour la région de l'Est. La plupart du temps, elle a été combinée à la ligne 23 seulement (ménages non familiaux d'une personne dont le soutien du ménage est une femme de 65 ans ou plus) qui avait une population trois fois plus importante que celle de la ligne 22. En outre, la figure 4 montre que la fraction de sondage de la ligne 22 était inférieure à 18% pour l'Est tandis qu'elle était supérieure à 19% à la ligne 23. Il n'est donc pas surprenant de constater que la ligne 22 présente un écart de - 2% pour l'Est.

La matrice des ménages comportait beaucoup moins de colonnes que de lignes (4 comparativement à 25). Par conséquent, il n'a fallu effectuer qu'un petit nombre de regroupements. De plus, les fractions de sondage étaient à peu près les mêmes pour chaque colonne. Conséquemment, les colonnes de la matrice des ménages montraient de très petits écarts au niveau régional (c'est-à-dire ne dépassant pas 0.13% pour toutes les colonnes de chaque région).

Toutes les lignes de la matrice des personnes avaient des écarts de moins de ± 0.1% pour l'ensemble des régions sauf les lignes 6, 7, 8, 9, 19, 20, 21 et 22. Toutes ces lignes présentaient des écarts de moins de ± 1% pour la totalité des régions sauf la ligne 7 pour l'Ontario où l'écart était inférieur à 3%. Ces divergences ont toutes été causées par la combinaison de lignes de faible valeur avec des lignes de valeur considérable ayant des fractions de sondage différentes. La ligne 6 a souvent été combinée avec la ligne 7, la ligne 8 avec la ligne 9, la ligne 19 avec la ligne 20 et la ligne 21 avec la ligne 22. La ligne 6 a une valeur beaucoup plus élevée que la ligne 7, la ligne 9 a aussi une valeur considérablement supérieure à celle de la ligne 8, la valeur de la ligne 19 est beaucoup plus importante que celle de la ligne 20 et la valeur de la ligne 22 surpasse nettement celle de la ligne 21.

Les valeurs des colonnes relatives aux autres langues maternelles de la matrice des personnes étaient généralement sous-estimées pour l'Est du pays (écarts négatifs de jusqu'à 9%) et, dans une moindre mesure, pour le Québec. Les valeurs des colonnes de la langue maternelle française tendaient à être surestimées pour l'Ouest (écarts positifs de jusqu'à 6%) et dans une moindre mesure pour l'Ontario. On a aussi constaté que les valeurs relatives à la langue maternelle anglaise avaient tendance à être surestimées pour le Québec (écarts positifs de moins de 2%). Ces résultats étaient attribuables au fait que les colonnes de la langue maternelle anglaise, de la langue maternelle française et des autres langues maternelles étaient habituellement toutes combinées pour l'Est et le Québec tandis que les colonnes relatives à la langue maternelle française et aux autres langues maternelles étaient généralement combinées pour l'Ouest et l'Ontario. La fraction de sondage relative à la langue maternelle française avait tendance à être la plus élevée, suivie de la fraction établie pour la langue maternelle anglaise, puis pour les autres langues maternelles. La fraction de sondage peu élevée déterminée pour les autres langues maternelles était attribuable en

Figure 1. Écarts selon les lignes de la matrice de pondération des ménages

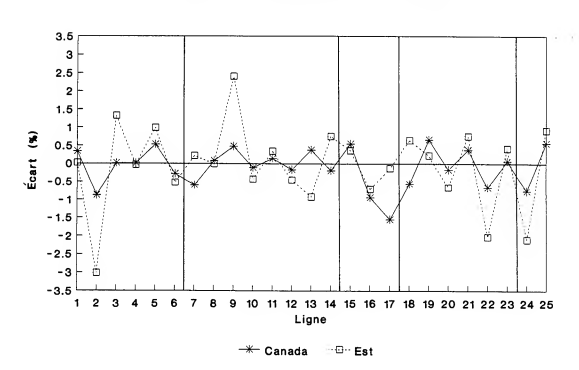


Figure 2. Pourcentage de la population par ligne de la matrice de pondération des ménages

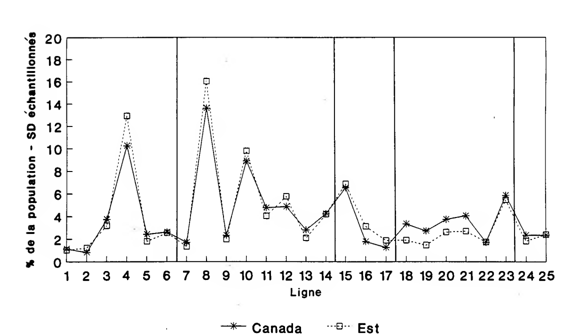


Figure 3. Pourcentage des RP pour lesquelles les lignes de la matrice de pondération des ménages ont ete combinées

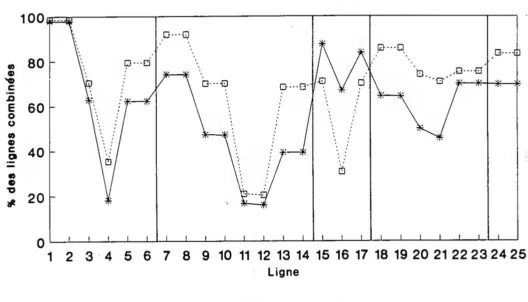
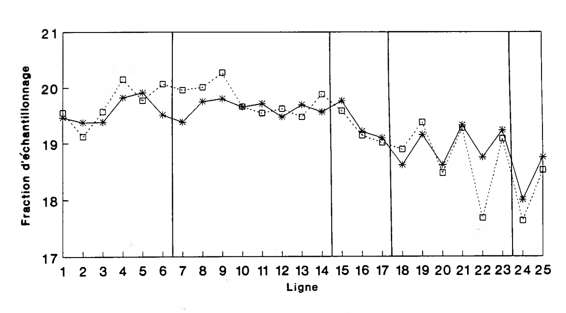


Figure 4. Fraction d'échantillonnage selon les lignes de la matrice de pondération des ménages



partie aux répondants qui avaient tendance à donner des réponses multiples (incluses dans la catégorie "Autres") plus souvent sur la formule 2A que sur la formule 2B (voir le chapitre VII).

En plus des regroupements, des écarts entre les estimations-échantillon et les chiffres de population peuvent survenir lorsque les résultats de l'application de la MIQ ne convergent pas. Un maximum de 80 itérations a été permis pour assurer la convergence des résultats, ce qui s'est avéré suffisant pour toutes les matrices des ménages; en fait, 98% d'entre elles ont nécessité moins de 20 itérations. Toutefois, le nombre moyen d'itérations pour les matrices des personnes était beaucoup plus élevé. En fait, il n'y avait toujours pas convergence pour 385 d'entre elles (7.2%) après 80 itérations. Cela touchait uniquement les colonnes étant donné que la MIQ se termine toujours par les lignes, de sorte que les écarts entre les lignes combinées sont toujours nuls. Il n'est pas nécessaire qu'il y ait convergence exacte pour interrompre l'application de la MIQ et en fait, un tel résultat est rarement atteint. Toutefois, on peut s'attendre à ce que les écarts au niveau des colonnes soient plus grands si la MIQ ne permet pas d'atteindre le niveau de convergence requis après 80 itérations. Le manque de convergence peut être causé par le manque de concordance des contraintes imposées aux lignes et aux colonnes. Les contraintes peuvent devenir incompatibles dans les cas où, après le regroupement de lignes et de colonnes, il y a un bloc de cellules dans la matrice pour lesquelles il n'y a pas d'unités échantillonnées, mais quelques unités dans la population. Cela peut créer une situation où il est impossible de rentrer les estimations-échantillon égales aux chiffres de population à la fois pour les lignes et les colonnes simultanément. Le fait d'utiliser l'âge pour définir tant les lignes que les colonnes de la matrice des personnes a effectivement donné lieu à des contraintes incompatibles dans au moins une matrice, ce qui a entraîné une surestimation des enfants de 0 à 14 ans dans les familles de recensement.

Pour plus de renseignements sur cette étude, voir Daoust et Bankier (1989).

VII. EXAMEN DE LA COHÉRENCE ENTRE LES ESTIMATIONS-ÉCHANTILLON ET LES CHIFFRES DE POPULATION

Les contraintes relatives à la taille qui sont imposées aux lignes et aux colonnes des matrices de recoupement et qui sont nécessaires à une application efficace de la MIQ, limitent le nombre de lignes et de colonnes que les matrices pourraient avoir. Conséquemment, plusieurs caractéristiques importantes ont été groupées au moment de la formation des lignes et des colonnes. Par suite de ce groupement, le niveau de cohérence entre les estimations-échantillon et les chiffres de population pour ces caractéristiques a été réduit. De plus, de nombreuses régions géographiques d'intérêt ne se composent pas toujours de RP complètes (voir le chapitre V, section A). Dans ces régions, la cohérence pour l'ensemble des caractéristiques dépend donc de la mesure dans laquelle les régions se composent de RP complètes.

L'examen de la cohérence porte sur les divergences entre les estimations-échantillon et les chiffres de population (exprimées en tant que pourcentages des chiffres de population) pour les mêmes caractéristiques de base que celles de l'examen du biais d'échantillonnage pour les régions géographiques suivantes:

(a) divisions de recensement;

(b) subdivisions de recensement;

(c) secteurs de recensement et secteurs de recensement provinciaux;

(d) secteurs de dénombrement.

L'examen de la cohérence a aussi porté sur:

(e) les cellules des matrices de pondération;

(f) la caractéristique "langue maternelle".

Comme il est indiqué dans le chapitre VI, section B, les écarts entre les estimations-échantillon et les chiffres de population ont été calculés de la façon suivante:

$$\text{écart} = \frac{\text{estimations-échantillon - chiffres de population}}{\text{chiffres de population}} \times 100$$

A. **Divisions de recensement (DR)**

Les centiles du tableau 3 résument les résultats relatifs au niveau de cohérence, pour l'ensemble des 260 DR échantillonnées au Canada, d'une grande variété de caractéristiques de base. Pour chaque caractéristique, N% des DR présentaient des écarts inférieurs au N° centile tandis que 100 - N% des DR avaient des divergences supérieures au N° centile. La divergence se situait donc entre le 10° et le 90° centile pour 80% des DR, entre le 25° et le 75° centile pour 50% des DR, etc. Par exemple, l'écart pour le groupe d'âge 0-5 ans variait entre -1.62% et 1.58% pour 80% des DR.

Toutes les DR sont formées uniquement de RP complètes. Ainsi, les caractéristiques représentées dans la matrice par une ligne ou une colonne qui n'était jamais ou presque jamais combinée à une autre présentaient une cohérence presque parfaite au niveau des DR[9]. Ces caractéristiques étaient le sexe, le groupe d'âge 65 ans et plus, toutes les caractéristiques des familles de recensement, les logements possédés, les logements loués, les ménages d'une personne et les ménages hors famille de recensement. Le niveau de cohérence observé pour les autres caractéristiques n'était pas parfait, mais demeurait quand même assez bon, sauf pour les caractéristiques représentant seulement un faible pourcentage de la population dans la plupart des DR, par exemple les immeubles à appartements de cinq étages ou plus et les ménages formés de plusieurs familles de recensement. Les représentations graphiques (ne figurant pas dans le présent rapport) des écarts par rapport aux chiffres de population montrent qu'en général, plus les chiffres de population d'une DR sont élevés, meilleure est la cohérence, et ce, pour toutes les caractéristiques.

La dernière colonne du tableau 3 donne le pourcentage de DR pour lesquelles l'estimation obtenue par la MIQ se rapprochait davantage des chiffres de population que l'estimation établie à l'aide d'un coefficient de pondération simple d'environ 5[10]. L'estimation obtenue par la MIQ était plus juste dans la majorité des DR pour l'ensemble des caractéristiques sauf les immeubles à appartements de cinq étages ou plus et les ménages formés de plusieurs familles de recensement.

---

9 Même lorsque les caractéristiques présentent une cohérence parfaite, les totalisations des caractéristiques de base établies à partir des données-échantillon ne concordent pas parfaitement avec les totalisations des mêmes caractéristiques basées sur les données intégrales. Cette différence est attribuable au fait que les résidents des logements collectifs qui n'avaient pas à répondre aux questions-échantillon (voir chapitre II, section B) sont inclus dans les totalisations faites à partir des données intégrales, mais exclus des totalisations fondées sur les données-échantillon.

10 Le coefficient de pondération simple (discuté ici et ailleurs également dans ce chapitre) pour chaque unité (personne ou logement) correspondait en fait à l'inverse de la fraction de sondage du ménage pour le SD dans lequel l'unité se trouvait.

Tableau 3. Centiles des écarts entre les estimations-échantillon et les chiffres de population des DR (en termes de pourcentages des chiffres de population), et pourcentage des DR où il y a réduction des écarts

Caractéristique	Centiles 10ième	25ième	50ième	75ième	90ième	Pourcentage des DR où les écarts obtenus par la MIQ était inférieur à ceux obtenus avec une pondération simple
Caractéristiques - Personnes						
Hommes	0.00	0.00	0.00	0.00	0.00	100
Femmes	0.00	0.00	0.00	0.00	0.00	100
Population totale	0.00	0.00	0.00	0.00	0.00	100
Âge, 0-5	- 1.62	- 0.57	0.06	0.93	1.58	81
Âge, 6-14	- 1.08	- 0.62	- 0.05	0.36	1.02	88
Âge, 15-24	- 0.46	- 0.20	0.00	0.16	0.41	94
Âge, 25-34	- 1.23	- 0.48	- 0.01	0.39	0.83	82
Âge, 35-44	- 0.90	- 0.44	0.03	0.51	1.05	86
Âge, 45-54	- 1.00	- 0.51	- 0.03	0.47	0.91	89
Âge, 55-64	- 1.06	- 0.40	0.09	0.52	1.23	88
Âge, 65 et plus	0.00	0.00	0.00	0.00	0.00	100
Célibataires	- 0.28	- 0.11	0.01	0.13	0.28	94
Personnes mariées	- 0.33	- 0.17	- 0.07	0.04	0.18	90
Veufs et veuves	- 4.49	- 2.01	- 0.07	2.14	5.26	63
Personnes divorcées	- 8.45	- 3.64	0.57	4.77	9.30	57
Personnes séparées	- 10.60	- 4.48	- 0.22	5.06	10.51	60
Caractéristiques - Familles de recensement						
Nombre total de familles	- 0.13	- 0.05	- 0.01	0.03	0.08	100
Familles époux-épouse	- 0.13	- 0.05	- 0.01	0.04	0.09	100
Familles monoparentales	- 0.13	- 0.07	- 0.04	0.00	0.04	100
Enfants dans les familles	0.08	- 0.01	0.05	0.10	0.20	100
Personnes dans les familles	- 0.00	0.00	0.00	0.01	0.02	100
Personnes hors famille	- 0.12	- 0.07	- 0.03	- 0.00	0.03	100
Caractéristiques - Ménages et logements						
Logements possédés	- 0.00	- 0.00	0.00	0.00	0.00	100
Logements loués	- 0.00	- 0.00	0.00	0.00	0.00	100
Maisons individuelles	- 0.53	- 0.21	0.03	0.22	0.48	74
App. dans un immeuble < que 5 étages	- 3.95	- 1.97	- 0.55	0.37	1.77	77
App. dans un immeuble ≥ que 5 étages	- 100.00	- 34.78	- 0.77	2.30	16.15	33
Logements mobiles	- 14.43	- 2.93	1.30	7.27	16.97	53
Autres logements	- 6.11	- 1.76	0.26	2.59	5.27	57
Ménages d'une personne	0.00	0.00	0.00	0.00	0.00	100
Ménages de 2 personnes	- 1.65	- 0.80	0.22	0.84	2.14	73
Ménages de 3 personnes	- 4.36	- 1.86	- 0.10	1.82	3.93	62
Ménages de 4 ou 5 personnes	- 2.73	- 1.11	0.06	1.23	2.62	63
Ménages de 6 personnes et plus	- 11.43	- 6.19	- 1.61	2.23	7.10	52
Ménages non familiaux	0.00	0.00	0.00	0.00	0.00	100
Ménages unifamiliaux	- 0.23	- 0.05	0.04	0.17	0.30	95
Ménages multifamiliaux	- 40.67	- 18.69	- 4.17	6.04	28.75	47
Soutiens du ménage < 25 ans	- 8.07	- 4.12	- 0.33	2.45	6.63	64
Soutiens du ménage 25-34 ans	- 1.76	- 0.73	0.12	0.84	1.59	75
Soutiens du ménage 35-44 ans	- 1.99	- 0.81	0.04	0.94	1.75	75
Soutiens du ménage 45-64 ans	- 1.33	- 0.56	0.11	0.66	1.27	79
Soutiens du ménage > 64 ans	- 1.44	- 0.61	- 0.02	0.38	1.05	87
Soutiens du ménage sexe masculin	- 0.75	- 0.38	- 0.09	0.18	0.41	73
Soutiens du ménage sexe féminin	- 1.57	- 0.56	0.24	1.16	2.64	73

B. Subdivisions de recensement (SDR)

Le tableau 4 résume les résultats concernant la cohérence entre les estimations-échantillon et les chiffres de population pour toutes les SDR échantillonnées au Canada dont les chiffres de population[11] sont supérieurs à 50. Il porte sur les mêmes caractéristiques que celles du tableau 3. Les SDR ne sont pas toujours uniquement formées de RP entières. En outre, elles sont beaucoup plus petites en moyenne que les DR. Par conséquent, le niveau de cohérence n'est pas aussi bon pour les SDR que pour les DR. Les estimations obtenues par la MIQ étaient meilleures que celles qui avaient été établies au moyen de coefficients de pondération simples pour la majorité des SDR et pour presque toutes les caractéristiques. De façon générale, comme dans le cas des DR, la cohérence devenait meilleure à mesure que les chiffres de population des SDR augmentaient, et ce, pour toutes les caractéristiques.

C. Secteurs de recensement (SR) et secteurs de recensement provinciaux (SRP)

Le tableau 5 résume les résultats relatifs au niveau de cohérence pour tous les SR échantillonnés au Canada dont les chiffres de population sont supérieurs à 50, et le tableau 6, pour tous les SRP échantillonnés au Canada dont les chiffres de population sont aussi supérieurs à 50. Les SR et les SRP sont en moyenne plus populeux que les SDR. Les SRP sont légèrement plus populeux en moyenne que les SR, mais les limites des SR ont été mieux respectées que celles des SRP au moment de la formation des RP. Conséquemment, la cohérence était supérieure au niveau des SR qu'au niveau des SRP pour la plupart des caractéristiques tandis qu'elle était meilleure au niveau des SRP qu'au niveau des SDR pour la majorité des caractéristiques. Les caractéristiques faisant exception à cette règle étaient généralement celles qui présentaient une cohérence médiocre à tous les niveaux géographiques. La cohérence des estimations obtenues par la MIQ était supérieure à celle des estimations faites à partir de coefficients de pondération simples pour la majorité des SR et des SRP, et ce, pour presque toutes les caractéristiques.

D. Secteurs de dénombrement (SD)

Les SD sont les composantes des RP. Tous les SD sauf cinq qui ont fait l'objet d'un traitement spécial (voir le chapitre VI, section A) ne constituaient qu'une partie d'une RP, ce qui représente le plus bas niveau auquel les estimations-échantillon sont forcées de concorder avec les chiffres de population. De plus, les coefficients de pondération initiaux utilisés étaient les mêmes pour toutes les personnes et tous les ménages de la même RP, même si la fraction de sondage différait entre les SD de la RP, tandis que les coefficients de pondération simples ont été calculés au niveau des SD. Par conséquent, on ne peut pas s'attendre à ce que la cohérence au niveau des SD soit aussi bonne qu'à des niveaux plus élevés. En fait, le tableau 7 montre que la cohérence des estimations établies par la MIQ était supérieure à celle des estimations faites à l'aide de coefficients de pondération simples

---

[11] On parle ici du chiffre de la population de la caractéristique à l'étude. Par exemple, le niveau de cohérence du groupe d'âge 0 - 5 ans est défini à partir de tous les SDR dans lesquelles il y avait plus de 50 personnes dans le groupe d'âge 0-5 ans. La même définition s'applique aux tableaux 5, 6, et 7.

Tableau 4. Centiles des écarts entre les estimations-échantillon et les chiffres de population des SDR (en termes de pourcentages des chiffres de population), et pourcentage des SDR où il y a réduction des écarts

Caractéristique	Centiles					Pourcentage des SDR où les écarts obtenus par la MIQ était inférieur à ceux obtenus avec une pondération simple
	10ième	25ième	50ième	75ième	90ième	
Caractéristiques - Personnes						
Hommes	- 9.15	- 3.46	0.00	3.26	9.75	60
Femmes	- 9.38	- 3.44	0.00	3.34	8.97	60
Population totale	- 7.56	- 2.71	0.00	2.59	7.93	58
Âge, 0-5	- 20.36	- 7.44	0.05	7.28	19.33	65
Âge, 6-14	- 20.42	- 7.16	- 0.16	7.14	19.61	65
Âge, 15-24	- 20.30	- 7.40	0.00	6.23	19.22	66
Âge, 25-34	- 17.54	- 6.53	- 0.05	6.29	17.55	62
Âge, 35-44	- 17.96	- 6.29	- 0.10	6.63	18.49	64
Âge, 45-54	- 19.38	- 7.09	- 0.11	6.47	19.73	65
Âge, 55-64	- 20.51	- 7.11	0.08	7.53	19.78	65
Âge, 65 et plus	- 19.39	- 6.41	0.00	7.05	19.12	67
Célibataires	- 13.63	- 4.81	- 0.02	4.60	13.08	62
Personnes mariées	- 8.50	- 3.11	- 0.06	3.07	8.71	57
Veufs et veuves	- 18.22	- 7.61	0.41	8.82	19.97	58
Personnes divorcées	- 21.57	- 8.48	0.61	10.55	20.73	54
Personnes séparées	- 22.91	- 10.45	0.30	9.52	20.13	56
Caractéristiques - Familles de recensement						
Nombre total de familles	- 7.15	- 2.59	- 0.02	2.78	7.48	56
Familles époux-épouse	- 8.19	- 2.85	- 0.02	2.95	8.18	59
Familles monoparentales	- 10.41	- 1.93	- 0.04	0.84	9.53	82
Enfants dans les familles	- 14.51	- 5.02	0.05	5.05	14.43	63
Personnes dans les familles	- 9.44	- 3.19	0.00	3.24	9.53	61
Personnes hors famille	- 19.36	- 6.80	- 0.03	5.89	18.96	68
Caractéristiques - Ménages et logements						
Logements possédés	- 6.75	- 2.51	0.00	2.41	6.80	54
Logements loués	- 13.97	- 3.79	0.00	3.68	13.58	70
Maisons individuelles	- 5.94	- 2.36	- 0.01	2.23	5.98	40
App. dans un immeuble < que 5 étages	- 8.04	- 3.38	- 0.51	2.07	7.25	64
App. dans un immeuble ≥ que 5 étages	- 6.44	- 2.37	0.12	1.93	6.88	36
Logements mobiles	- 12.25	- 5.52	0.80	6.84	15.26	48
Autres logements	- 13.81	- 5.24	0.09	6.08	14.79	55
Ménages d'une personne	- 15.74	- 4.22	0.00	4.73	15.46	72
Ménages de 2 personnes	- 15.78	- 5.99	0.11	6.57	16.36	57
Ménages de 3 personnes	- 20.29	- 8.25	0.02	8.85	20.53	51
Ménages de 4 ou 5 personnes	- 15.91	- 6.47	0.16	7.29	15.98	51
Ménages de 6 personnes et plus	- 25.23	- 12.60	- 2.12	8.48	22.15	54
Ménages non familiaux	- 15.97	- 5.15	0.00	5.14	16.05	69
Ménages unifamiliaux	- 7.10	- 2.58	0.07	2.95	7.41	56
Ménages multifamiliaux	- 26.53	- 13.34	- 3.13	8.55	21.56	39
Soutiens du ménage < 25 ans	- 19.76	- 9.84	- 1.42	6.33	16.59	59
Soutiens du ménage 25-34 ans	- 16.94	- 6.20	0.03	5.73	15.21	61
Soutiens du ménage 35-44 ans	- 15.77	- 5.70	0.12	5.50	15.01	63
Soutiens du ménage 45-64 ans	- 15.44	- 5.68	0.04	5.78	14.76	58
Soutiens du ménage > 64 ans	- 15.74	- 5.77	0.02	5.66	15.70	65
Soutiens du ménage sexe masculin	- 7.63	- 2.94	- 0.03	2.71	7.20	48
Soutiens du ménage sexe féminin	- 15.84	- 5.72	0.12	6.66	18.07	59

Tableau 5. Centiles des écarts entre les estimations-échantillon et les chiffres de population des SR (en termes de pourcentages des chiffres de population), et pourcentage des SR où il y a réduction des écarts

Caractéristique	Centiles					Pourcentage des SR où les écarts obtenus par la MIQ était inférieur à ceux obtenus avec une pondération simple
	10ième	25ième	50ième	75ième	90ième	
Caractéristiques - Personnes						
Hommes	- 2.13	0.00	0.00	0.00	1.93	83
Femmes	- 1.95	0.00	0.00	0.00	1.93	82
Population totale	- 1.67	0.00	0.00	0.00	1.52	81
Âge, 0-5	- 7.69	- 3.31	0.01	3.52	8.43	76
Âge, 6-14	- 6.95	- 2.51	- 0.02	2.44	6.69	77
Âge, 15-24	- 4.85	- 1.15	0.00	1.00	4.66	83
Âge, 25-34	- 5.28	- 2.27	- 0.02	2.12	5.57	73
Âge, 35-44	- 5.64	- 2.14	- 0.09	2.04	5.46	74
Âge, 45-54	- 6.28	- 2.00	- 0.12	2.01	5.84	77
Âge, 55-64	- 5.64	- 2.00	0.00	2.22	6.46	79
Âge, 65 et plus	- 5.17	0.00	0.00	0.00	5.27	85
Célibataires	- 2.85	- 0.60	- 0.04	0.63	2.82	81
Personnes mariées	- 2.32	- 0.74	- 0.07	0.60	2.01	77
Veufs et veuves	- 16.22	- 7.93	0.31	8.20	17.01	59
Personnes divorcées	- 21.36	- 10.75	0.36	11.23	21.66	55
Personnes séparées	- 24.42	- 12.65	0.17	11.96	24.32	57
Caractéristiques - Familles de recensement						
Nombre total de familles	- 2.03	- 0.30	0.00	0.23	1.70	79
Familles époux-épouse	- 2.22	- 0.33	0.01	0.25	2.05	81
Familles monoparentales	- 5.68	- 0.26	- 0.01	0.17	6.13	86
Enfants dans les familles	- 3.28	- 0.34	0.03	0.52	3.39	84
Personnes dans les familles	- 2.07	- 0.03	0.00	0.06	1.88	84
Personnes hors famille	- 4.37	- 0.29	- 0.02	0.13	4.26	84
Caractéristiques - Ménages et logements						
Logements possédés	- 1.81	- 0.01	0.00	0.01	2.02	82
Logements loués	- 2.87	- 0.01	0.00	0.01	2.73	81
Maisons individuelles	- 2.90	- 1.14	0.04	1.08	2.96	58
App. dans un immeuble < 5 étages	- 7.43	- 3.19	- 0.42	1.95	6.14	52
App. dans un immeuble ≥ 5 étages	- 7.16	- 2.74	0.01	2.85	8.24	39
Logements mobiles	- 9.78	- 4.81	0.41	5.20	12.27	46
Autres logements	- 9.80	- 4.19	0.40	4.76	11.29	52
Ménages d'une personne	- 4.59	0.00	0.00	0.00	4.35	85
Ménages de 2 personnes	- 6.91	- 2.82	0.30	3.55	7.69	66
Ménages de 3 personnes	- 14.29	- 7.20	- 0.41	6.58	13.31	54
Ménages de 4 ou 5 personnes	- 11.01	- 5.34	0.06	5.07	10.44	56
Ménages de 6 personnes et plus	- 31.23	- 18.37	- 4.06	11.43	26.77	51
Ménages non familiaux	- 4.34	0.00	0.00	0.00	3.92	84
Ménages unifamiliaux	- 2.10	- 0.67	0.08	0.73	2.08	73
Ménages multifamiliaux	- 35.67	- 21.14	- 4.87	10.28	26.72	46
Soutiens du ménage < 25 ans	- 22.02	- 11.03	- 1.15	9.01	20.36	57
Soutiens du ménage 25-34 ans	- 7.67	- 3.45	0.08	3.59	7.83	68
Soutiens du ménage 35-44 ans	- 8.45	- 3.82	0.02	3.68	8.15	67
Soutiens du ménage 45-64 ans	- 6.34	- 3.01	0.07	3.03	6.34	67
Soutiens du ménage > 64 ans	- 7.95	- 2.87	- 0.12	2.85	7.70	74
Soutiens du ménage sexe masculin	- 2.89	- 1.06	0.00	1.00	2.53	68
Soutiens du ménage sexe féminin	- 6.77	- 2.48	0.00	2.59	7.09	70

Tableau 6. Centiles des écarts entre les estimations-échantillon et les chiffres de population des SRP (en termes de pourcentages des chiffres de population), et pourcentage des SRP où il y a réduction des écarts

Caractéristique	Centiles 10ième	25ième	50ième	75ième	90ième	Pourcentage des SRP où les écarts obtenus par la MIQ était inférieur à ceux obtenu avec une pondération simple
Caractéristiques - Personnes						
Hommes	- 2.56	- 1.03	0.00	1.03	2.67	70
Femmes	- 2.71	- 0.98	0.00	1.05	2.60	71
Population total	- 2.11	- 0.81	0.00	0.76	2.06	70
Âge, 0-5	- 9.10	- 3.88	0.02	4.24	8.75	70
Âge, 6-14	- 7.30	- 3.34	- 0.11	3.33	7.34	71
Âge, 15-24	- 6.27	- 2.68	0.00	2.52	6.69	73
Âge, 25-34	- 6.18	- 2.85	- 0.06	2.76	6.29	67
Âge, 35-44	- 6.96	- 2.92	0.01	3.08	6.51	70
Âge, 45-54	- 8.10	- 3.52	0.05	3.54	8.00	70
Âge, 55-64	- 7.73	- 3.34	0.06	3.47	8.04	69
Âge, 65 et plus	- 7.67	- 2.79	0.00	2.75	7.08	75
Célibataires	- 3.86	- 1.42	0.02	1.58	3.91	69
Personnes mariées	- 2.51	- 1.04	- 0.09	0.95	2.41	70
Veufs et veuves	- 14.31	- 7.61	- 0.13	7.21	14.74	59
Personnes divorcées	- 22.77	- 11.34	1.89	13.72	25.20	53
Personnes séparées	- 25.88	- 14.01	- 1.05	13.12	26.44	54
Caractéristiques - Familles de recensement						
Nombre total de familles	- 2.06	- 0.84	0.00	0.84	2.18	70
Familles époux-épouse	- 2.40	- 0.92	0.00	0.97	2.43	72
Familles monoparentales	- 11.62	- 4.51	- 0.05	4.20	10.81	74
Enfants dans les familles	- 4.25	- 1.55	0.07	1.77	4.23	72
Personnes dans les familles	- 2.48	- 0.98	0.00	0.95	2.58	74
Personnes hors famille	- 7.23	- 2.61	0.00	2.63	7.10	77
Caractéristiques - Ménages et logements						
Logements possédés	- 2.18	- 0.87	- 0.00	0.76	2.07	67
Logements loués	- 6.58	- 2.32	0.00	2.34	6.37	72
Maisons individuelles	- 2.35	- 1.17	0.02	1.16	2.44	51
App. dans un immeuble < 5 étages	- 8.73	- 3.96	- 0.46	2.77	7.15	62
App. dans un immeuble ≥ 5 étages	- 11.49	- 5.05	0.79	4.40	9.71	40
Logements mobiles	- 17.70	- 8.38	0.59	7.61	16.63	48
Autres logements	- 14.48	- 6.59	0.35	6.75	15.89	55
Ménages d'une personne	- 8.38	- 2.97	0.00	2.90	8.04	75
Ménages de 2 personnes	- 7.94	- 3.73	0.17	3.72	8.00	62
Ménages de 3 personnes	- 14.18	- 7.20	- 0.00	6.97	14.41	54
Ménages de 4 ou 5 personnes	- 8.82	- 4.52	0.21	4.72	9.09	56
Ménages de 6 personnes et plus	- 27.93	- 14.93	- 1.40	12.04	27.55	51
Ménages non familiaux	- 8.14	- 2.63	0.00	2.92	7.05	75
Ménages unifamiliaux	- 2.20	- 0.93	0.06	1.01	2.29	68
Ménages multifamiliaux	- 40.55	- 30.78	- 9.42	14.13	29.67	59
Soutiens du ménage < 25 ans	- 22.45	- 11.77	- 1.14	10.04	20.82	56
Soutiens du ménage 25-34 ans	- 8.15	- 3.71	0.12	4.15	8.50	64
Soutiens du ménage 35-44 ans	- 8.76	- 4.22	0.11	4.49	9.25	63
Soutiens du ménage 45-64 ans	- 6.77	- 3.23	0.19	3.41	6.76	63
Soutiens du ménage > 64 ans	- 8.77	- 3.60	- 0.13	3.30	7.93	69
Soutiens du ménage sexe masculin	- 2.77	- 1.39	- 0.04	1.16	2.54	59
Soutiens du ménage sexe féminin	- 8.33	- 3.49	0.11	3.98	9.07	64

Tableau 7. Pourcentage de tous les SD échantillonés au Canada dont le chiffre de population excède 50 et où les écarts obtenus par la MIQ étaient inférieurs à ceux obtenus à partir des coefficients de pondération simples

Caractéristiques	Pourcentage de SD où il y a eu réduction des écarts	Caractéristiques	Pourcentage de SD où il y a eu réduction des écarts
Hommes	7	Logements possédés	5
Femmes	7	Logements loués	4
Âge, 0-5	19	App. dans un immeuble < 5 étages	2
Âge, 6-14	17	App. dans un immeuble ≥ 5 étages	1
Âge, 15-24	17	Logements mobiles	0
Âge, 25-34	15	Autres logements	3
Âge, 35-44	16	Ménages d'une personne	9
Âge, 45-54	19	Ménages de 2 personnes	16
Âge, 55-64	19	Ménages de 3 personnes	18
Âge, 65 et plus	15	Ménages de 4 ou 5 personnes	12
Célibataires	10	Ménages de 6 personnes et plus	3
Personnes mariées	7	Ménages non familiaux	9
Veufs et veuves	11	Ménages unifamiliaux	6
Personnes divorcées	10	Ménages multifamiliaux	*
Personnes séparées	6	Soutiens du ménage < 25 ans	5
Nombre total de familles	6	Soutiens du ménage 25-34 ans	16
Familles époux-épouse	7	Soutiens du ménage 35-44 ans	16
Familles monoparentales	8	Soutiens du ménage 45-64 ans	16
Enfants dans les familles	11	Soutiens du ménage > 64 ans	11
Personnes dans les familles	7	Soutiens du ménage sexe masculin	7
Personnes hors famille	14	Soutiens du ménage sexe féminin	12

* Aucun SD ne comptait plus de 50 ménages multifamiliaux.

pour moins de 20% de tous les SD échantillonnés au Canada dont les chiffres de population excèdent 50, pour toutes les caractéristiques observées.

E. Cellules des matrices de pondération

La MIQ permet seulement de garantir que les totaux estimatifs des lignes et des colonnes de la matrice de recoupement concordent avec les chiffres de population correspondants. Il n'y a aucun mécanisme de contrôle au niveau des cellules individuelles de la matrice. On a étudié la cohérence au niveau des cellules pour les matrices des personnes et des ménages pour cinq RP choisies au hasard (seules les cellules comportant des unités échantillonnées étaient incluses). Sur les dix matrices, la cohérence des estimations par la MIQ était meilleure que celle des estimations simples utilisant des coefficients de pondération égaux à la fraction de sondage inverse des ménages du RP pour 58% des cellules et moins bonne pour 42% d'entre elles. Les cellules pour lesquelles les estimations obtenues par la MIQ étaient supérieures aux estimations simples avaient en moyenne des chiffres de population plus élevés que ceux des cellules restantes. De plus, les écarts avaient tendance à décroître à mesure que les chiffres de population des cellules augmentaient.

On a constaté que l'application de la MIQ entraînait une nette surestimation des valeurs des cellules. Sur les dix matrices, la valeur des cellules a été surestimée dans 58% des cas et sous-estimée dans 42% des cas. Aucune unité n'a été choisie dans l'échantillon pour 22% des cellules n'ayant pas des chiffres de population nuls. Étant donné que les estimations des lignes et des colonnes doivent concorder avec les chiffres de population, cela signifie que les estimations des autres cellules doivent être augmentées afin de contrebalancer la sous-estimation de ces cellules. C'est probablement ce qui explique en grande partie sinon entièrement la surestimation. Toutefois, étant donné que la taille des cellules sans unités échantillonnées est petite, la surestimation est également faible.

Pour plus de renseignements sur l'examen de la cohérence, voir Rathwell (1990).

F. Cohérence de la caractéristique "langue maternelle"

Une étude distincte a été faite sur la cohérence des réponses à la question sur la langue maternelle. A l'échelle nationale, la divergence entre les estimations-échantillon et les chiffres de population pour les catégories de réponse "anglais seulement" et "français seulement" était infime. Toutefois, des écarts plus importants ont été relevés dans les cas de réponses multiples (- 9.67%), particulièrement pour les catégories "anglais et français" (- 7.35%) et "anglais et langues non officielles" (- 11.64%).

Un pourcentage plus élevé de personnes ont donné des réponses multiples sur la formule 2A (3.9% avant l'imputation) que sur la formule 2B (3.3% avant l'imputation) de sorte que la fraction de sondage pour cette catégorie était faible. De plus, les réponses multiples ont été incluses dans les colonnes "autre langue maternelle" de la matrice de pondération des personnes qui ont souvent été combinées avec les colonnes "anglais seulement" et "français seulement". Ces deux facteurs ajoutés au fait que les réponses multiples représentaient un pourcentage relativement faible de la population ont donné lieu à une importante sous-estimation du nombre de personnes ayant plus d'une langue maternelle (voir le chapitre VI, section B).

Le pourcentage plus élevé de réponses multiples sur la formule 2A que sur la formule 2B ne peut être imputé au contrôle et à l'imputation ou au traitement des données pas plus qu'à la variabilité de l'échantillonnage. Il semble que les répondants ont interprété la question sur la langue maternelle différemment sur la formule 2B que sur la formule 2A. Bien que la raison précise de cette différence d'interprétation soit inconnue, il est possible que les questions additionnelles sur la langue et l'origine ethnique de la formule 2B puissent avoir aidé à réduire le nombre de personnes ayant indiqué plus d'une langue maternelle en leur fournissant une autre occasion de déclarer toute autre langue parlée et (ou) origine. De plus, le fait que les instructions se rapportant aux questions, qui précisaient que les réponses multiples étaient permises, figuraient sur la formule 2A et non dans un livret distinct, comme pour la formule 2B, peut avoir eu une incidence sur la différence dans la fréquence des réponses multiples.

Pour plus de renseignements, voir Daoust (1988).

VIII. VARIANCE D'ÉCHANTILLONNAGE

L'erreur d'échantillonnage a deux composantes: la variance et le biais. La variance mesure la variabilité de l'estimation des valeurs que prendrait une variable par rapport à sa valeur moyenne obtenue pour des répétitions hypothétiques du processus d'enquête. Le biais est la différence entre la valeur moyenne que prendrait une estimation dans l'hypothèse d'une répétition de la procédure d'enquête et la valeur réelle de la variable observée. Le chapitre V, qui contient les résultats de l'étude du biais d'échantillonnage, décrit la nature et l'importance du biais dans les estimations-échantillon avant la pondération. Les chapitres VI et VII examinent le biais d'échantillonnage après la pondération. Même avec une méthode d'échantillonnage totalement dépourvue de biais, les résultats seraient sujets à la variance du simple fait que les estimations ont été produites à partir d'un échantillon. La variance peut être estimée à l'aide des données recueillies dans l'enquête-échantillon.[12] L'étude de la variance d'échantillonnage a été effectué afin d'estimer l'effet des méthodes d'échantillonnage et d'estimation sur les chiffres du recensement établis à partir de données-échantillon.

Statistique Canada produit des milliers de tableaux à partir des données-échantillon (c'est-à-dire celles recueillies sur la formule 2B). Théoriquement, on pourrait fournir une mesure de précision (à savoir l'estimation de la variance d'échantillonnage) pour chaque estimation présentée dans ces tableaux. Cette mesure tiendrait compte à la fois du plan de sondage et de la méthode d'estimation. En pratique cependant, il est impossible d'établir de telles mesures pour toutes les estimations du recensement en raison du coût élevé du traitement des données. On calcule donc la variance d'échantillonnage pour un sous-ensemble seulement des estimations du recensement. Cela permet ensuite d'évaluer les effets combinés du plan de sondage et de la méthode d'estimation sur la variance d'échantillonnage. On peut ensuite ajuster les estimations simples de la variance d'échantillonnage, qui sont peu coûteuses à produire, pour tenir compte de cet effet et produire une estimation de la variance d'échantillonnage pour n'importe quelle estimation du recensement.

La racine carrée de la variance d'échantillonnage, soit l'erreur-type, peut être estimée à l'aide des données présentées dans les tableaux 8 et 9. Le tableau 8 donne des estimations non ajustées (simples) de l'erreur-type d'estimations-échantillon. Les valeurs présentées dans ce tableau ont été établies en supposant que la fraction de sondage était de 1/5 dans un échantillon aléatoire simple et qu'on avait utilisé une pondération simple de 5. Les erreurs-types présentées dans le tableau 8 sont fonction à la fois de l'importance du "total estimé" et de la population de la région. Par exemple, pour une estimation de 250 personnes dans une région comptant 1 000 personnes, l'erreur-type non ajustée est de 25. Le tableau 8 contient les erreurs-types associées à un nombre restreint de valeurs du "total estimé" et du nombre total de personnes, de ménages, de logements ou de familles dans la région. La formule ci-après peut être utilisée pour calculer l'erreur-type non ajustée de n'importe quel total estimé, pour toute région, quelle qu'en soit l'importance:

---

[12] Malheureusement, la variance d'échantillonnage ne nous fournit aucune indication de l'importance de l'erreur non due à l'échantillonnage.

$$ETNA=\sqrt{\frac{4E(N-E)}{N}}$$

où ETNA est l'erreur-type non ajustée, E le total estimé et N le nombre total de personnes, de ménages, de logements, ou de familles dans la région. Par exemple, pour un total estimé de 750 personnes dans une région comptant 9 000 personnes, l'erreur-type non ajustée serait:

$$\sqrt{\frac{4(750)(9000-750)}{9000}}=52$$

Le tableau 9 contient les facteurs d'ajustement[13] par lesquels on doit multiplier les erreurs-types non ajustées pour tenir compte des effets combinés du plan de sondage et de la méthode d'estimation. Pour calculer ces facteurs d'ajustement, un échantillon de 401 RP (sur un total de 5 341) a été tiré. L'échantillon a été réparti parmi les dix provinces[14] de manière à obtenir de bonnes estimations de la variance d'échantillonnage au niveau provincial sans pour autant compromettre indûment la qualité des estimations au niveau national. Pour chaque RP dans l'échantillon, les estimations de la variance d'échantillonnage pour les estimations produites par la MIQ ont été calculées pour différentes sous-catégories de toutes les caractéristiques[15] énumérées dans le tableau 9. Les estimations de la variance d'échantillonnage aux niveaux provincial et national ont été obtenues en pondérant les estimations au niveau de la RP. Les facteurs d'ajustement pour chaque sous-catégorie de chacune des caractéristiques ont été calculés en divisant la racine carrée de ces estimations par l'erreur-type non ajustée correspondante. Des facteurs d'ajustement ont été calculés aux niveaux provincial et national pour chaque caractéristique en faisant la moyenne des facteurs d'ajustement de toutes les sous-catégories. Pour plus de renseignements sur la façon dont ces facteurs d'ajustement ont été calculés, voir Béland (1990).

Pour estimer l'erreur-type d'une estimation-échantillon donnée, l'utilisateur doit se reporter au tableau 9 pour obtenir le facteur d'ajustement s'appliquant à la caractéristique en question et multiplier ce facteur par l'erreur-type non ajustée correspondante dans le tableau 8. Lorsqu'une caractéristique ne figure pas au tableau 9, l'utilisateur doit se servir du facteur d'ajustement indiqué à la catégorie "Autres". Pour chaque caractéristique énumérée, le tableau 9 fournit les facteurs d'ajustement applicables aux niveaux national et provincial ainsi

---

13 On appelle communément "effet du plan" le carré des facteurs d'ajustement.

14 Le Yukon et les Territoires du Nord-Ouest ont été groupés avec la Colombie-Britannique.

15 Par exemple, la tranche $15 000 - $25 000 est l'une des catégories pour lesquelles une estimation de la variance d'échantillonnage a été calculée pour la caractéristique "revenu de la famille".

Tableau 8. Estimations non ajustées de l'erreur-type des estimations-échantillon

Total estimé	Nombre total de personnes, ménages, logements, où familles dans la région								
	500	1 000	2 500	5 000	10 000	25 000	50 000	100 000	250 000
50	15	15	15	15	15	15	15	15	15
100	18	19	20	20	20	20	20	20	20
250	22	25	30	30	30	30	30	30	30
500	0	30	40	40	45	45	45	45	45
1 000		0	50	55	60	60	65	65	65
2 500			0	70	85	95	95	100	100
5 000				0	100	130	130	140	140
10 000					0	150	180	190	200
25 000						0	220	270	300
50 000							0	320	400
100 000								0	490
250 000									0

Total estimé	Nombre total de personnes, ménages, logements, où familles dans la région					
	500 000	1 000 000	2 500 000	5 000 000	10 000 000	25 000 000
50	15	15	15	15	15	15
100	20	20	20	20	20	20
250	30	30	30	30	30	30
500	45	45	45	45	45	45
1 000	65	65	65	65	65	65
2 500	100	100	100	100	100	100
5 000	140	140	140	140	140	140
10 000	200	200	200	200	200	200
25 000	310	310	310	320	320	320
50 000	420	440	440	440	450	450
100 000	570	600	620	630	630	630
250 000	710	870	950	970	990	990
500 000	0	1 000	1 260	1 340	1 380	1 400
1 000 000		0	1 550	1 790	1 900	1 960
2 500 000			0	2 240	2 740	3 000
5 000 000				0	3 160	4 000
10 000 000					0	4 900

Tableau 9. Facteurs d'ajustement de l'erreur-type au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération (RP)

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP						
		1	50	75	90	95	99	100
Caractéristiques de la population								
Âge								
. Groupes d'âge 0-4, 5-9, 10-14, 15-19, 20-24, 25-29	0.18	0.05	0.19	0.29	0.35	0.49	0.52	0.60
. Groupes d'âge 30-34, 35-44, 45-54, 55-59, 60-64, 5+, 15+	0.36	0.13	0.33	0.46	0.51	0.56	0.61	0.74
. Groupe d'âge 65+	0.00	-	-	-	-	-	-	-
Sexe	0.00	-	-	-	-	-	-	-
État matrimonial								
. Célibataire, marié (sauf séparé(e))	0.25	0.04	0.23	0.31	0.42	0.49	0.55	0.68
. Séparé(e), divorcé(e), veuf (veuve)	0.88	0.55	0.84	0.98	1.06	1.15	1.20	1.33
Plus haut niveau de scolarité atteint/plus haut grade, certificat, ou diplôme obtenu/nombre total d'années de scolarité	0.90	0.75	0.95	1.06	1.14	1.19	1.25	1.38
Principal domaine d'études	1.20	0.84	1.16	1.22	1.28	1.35	1.43	1.51
Mobilité								
. N'ayant pas déménagé	1.21	0.83	1.23	1.27	1.32	1.36	1.41	1.58
. Ayant déménagé (migrants,non-migrants)	1.61	0.90	1.60	1.75	1.85	1.97	2.09	2.21
Période d'immigration								
. Avant 1946, 1946-1966	0.98	0.76	1.02	1.10	1.22	1.37	1.45	1.62
. 1967-1977, 1978-1982, 1983-1986	1.51	0.80	1.45	1.55	1.78	1.90	2.11	2.20
Âge à l'immigration	1.10	0.71	1.15	1.29	1.38	1.44	1.54	1.67
Lieu de naissance								
. Au Canada	1.09	0.82	1.08	1.16	1.18	1.20	1.21	1.33
. À l'étranger	1.35	1.11	1.34	1.43	1.60	1.67	1.75	1.91
Population immigrante/non immigrante	1.12	0.81	1.10	1.24	1.38	1.46	1.52	1.69
Citoyenneté								
. Du Canada, par naissance	1.13	0.88	1.14	1.17	1.20	1.27	1.32	1.58
. D'un autre pays	1.59	1.04	1.40	1.65	1.88	1.95	2.12	2.30

Tableau 9. Facteurs d'ajustement de l'erreur-type au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération (RP) - suite

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP						
		1	50	75	90	95	99	100
Origine ethnique								
. Français, anglais	1.20	0.73	1.16	1.25	1.31	1.40	1.46	1.65
. Autres	1.65	1.07	1.57	1.70	1.89	1.99	2.11	2.45
Langue parlée à la maison								
. Français, anglais, français et anglais, anglais et autres	1.12	0.50	1.09	1.35	1.75	1.89	2.09	2.19
. Autres	1.76	0.99	1.68	1.89	2.01	2.20	2.41	2.66
Langue officielle								
. Français, anglais, français et anglais	1.05	0.69	1.01	1.18	1.31	1.42	1.58	1.75
. Autres	1.49	0.90	1.50	1.68	1.76	1.79	1.91	2.01
Langue maternelle - anglais								
. Terre-Neuve, Ile-du-Prince-Édouard, Nouvelle-Écosse, Colombie-Britannique	0.92	0.24	0.96	1.45	1.62	1.90	2.23	2.45
. Québec	1.15	0.18	1.10	1.51	1.76	1.81	1.99	2.21
. Autres provinces	0.45	0.12	0.48	0.71	0.96	1.12	1.38	1.68
. Canada	0.53	-	-	-	-	-	-	-
Langue maternelle - français								
. Québec	0.42	0.14	0.45	0.52	0.61	0.76	0.91	1.19
. Nouveau-Brunswick	0.75	0.19	0.79	0.98	1.24	1.60	1.84	2.04
. Autres provinces	1.04	0.09	1.12	1.49	1.71	1.89	2.06	2.40
. Canada	0.77	-	-	-	-	-	-	-
Langue maternelle - autre	1.70	0.73	1.63	2.11	2.44	2.51	2.60	2.70
Industrie/profession	0.92	0.25	0.80	1.13	1.25	1.31	1.38	1.67
Travail en 1985	0.89	0.62	0.92	1.14	1.22	1.29	1.31	1.45
Semaines travaillées en 1985	0.94	0.68	0.99	1.18	1.29	1.33	1.39	1.69
Heures travaillées durant la semaine de référence	0.83	0.63	0.85	1.01	1.14	1.19	1.24	1.51
Date du dernier travail								
. En 1986, en 1985, avant 1985	0.89	0.60	0.94	0.99	1.05	1.11	1.20	1.33
. N'a jamais travaillé	1.18	0.80	1.15	1.34	1.43	1.50	1.67	1.82

Tableau 9. Facteurs d'ajustement de l'erreur-type au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération (RP) - suite

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP 1	50	75	90	95	99	100
Catégorie de travailleur								
. Travailleur rémunéré	0.72	0.56	0.75	0.86	0.93	0.95	0.98	1.09
. Travailleur autonome (entreprise ou ferme non constituée en société, travailleur familial non rémunéré)	0.93	0.68	0.96	1.08	1.13	1.15	1.18	1.31
Activité								
. Occupé	0.75	0.59	0.76	0.83	0.86	0.91	0.93	1.04
. En chômage	1.06	0.76	1.04	1.14	1.20	1.27	1.38	1.53
. Inactif	1.25	0.91	1.30	1.43	1.50	1.58	1.63	1.84
Principal source de revenu								
. Salaires et traitements	0.65	0.42	0.67	0.80	0.85	0.87	0.92	0.99
. Autres	1.05	0.71	1.00	1.12	1.17	1.20	1.24	1.48
Incapacité								
. Limité à la maison, à l'école, au travail	0.94	0.69	0.96	1.11	1.29	1.34	1.42	1.69
. Pas limité	0.61	0.41	0.58	0.69	0.74	0.78	0.81	0.84
Situation dans la famille de recensement								
. Époux, épouse, enfant	0.20	0.05	0.20	0.24	0.26	0.28	0.31	0.34
. Parent seul de sexe féminin	0.45	0.14	0.43	0.51	0.55	0.61	0.68	0.81
. Parent seul de sexe masculin, personne hors famille	0.68	0.35	0.65	0.79	0.89	0.99	1.14	1.32
Situation dans la famille économique								
. Époux, épouse	0.14	0.06	0.16	0.21	0.28	0.34	0.36	0.42
. Parent seul, enfant	0.32	0.16	0.34	0.39	0.44	0.47	0.53	0.68
. Autres membres de la famille	0.74	0.24	0.70	0.84	1.03	1.09	1.18	1.31
Nombre de personnes dans la famille de recensement	0.04	0.00	0.00	0.05	0.07	0.09	0.11	0.13
Nombre de personnes dans la famille économique	0.18	0.08	0.19	0.24	0.33	0.41	0.45	0.71
Âge de l'époux, de l'épouse, ou de la personne repère de la famille économique	1.42	0.80	1.37	1.53	1.60	1.78	1.91	2.08
Autres caractéristiques de le population	1.00	-	-	-	-	-	-	-

Tableau 9. Facteurs d'ajustement de l'erreur-type au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération (RP) - suite

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP						
		1	50	75	90	95	99	100
Caractéristiques du ménage et du logement								
Type de construction								
. Maisons individuelles	0.33	0.05	0.35	0.55	0.67	0.75	0.89	1.08
. App. dans un immeuble < 5 étages	0.57	0.12	0.56	0.70	0.83	0.99	1.26	1.44
. Autres	0.91	0.18	0.88	0.99	1.18	1.23	1.32	1.51
Mode d'occupation	0.00	-	-	-	-	-	-	-
Période de construction	0.78	0.61	0.75	0.82	0.89	0.99	1.24	1.49
Système de chauffage principal/ principal combustible pour le chauffage	0.87	0.18	0.86	1.04	1.12	1.25	1.32	1.47
Système de chauffage central								
. Avec chauffage central	0.42	0.09	0.38	0.54	0.60	0.70	0.89	1.19
. Sans chauffage central	0.78	0.23	0.79	0.91	1.03	1.12	1.20	1.39
Taille du ménage								
. Une personne	0.00	-	-	-	-	-	-	-
. Autres	0.76	0.19	0.72	1.09	1.17	1.21	1.30	1.53
Nombre de pièces	0.80	0.57	0.78	0.90	0.97	1.10	1.20	1.44
Âge du soutien de ménage								
. 25-34, 55-64, 65-74, 75+	0.25	0.06	0.24	0.35	0.48	0.53	0.62	0.94
. 0-24, 35-44, 45-54	0.92	0.38	0.90	1.05	1.14	1.21	1.30	1.49
Sexe du soutien de ménage								
. Homme	0.20	0.09	0.24	0.31	0.34	0.36	0.37	0.42
. Femme	0.47	0.16	0.43	0.54	0.64	0.74	0.89	1.09
Loyer brut/pourcentage du loyer brut dans le revenu du ménage	0.75	0.48	0.79	0.91	0.94	0.96	1.01	1.21
Principales dépenses de propriété/pourcentage des principales dépenses de propriété dans le revenu du ménage	0.84	0.62	0.87	0.95	1.01	1.04	1.11	1.29
Revenu du ménage	0.75	0.51	0.73	0.82	0.90	0.95	1.03	1.17
Valeur du logement	0.90	0.67	0.91	1.00	1.05	1.12	1.18	1.32
Copropriété enregistrée								
. Enregistrée	0.63	0.18	0.59	0.84	0.93	1.11	1.30	1.48
. Non enregistrée	0.15	0.07	0.14	0.19	0.28	0.39	0.47	0.59

Tableau 9. Facteurs d'ajustement de l'erreur-type au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération (RP) - suite

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP						
		1	50	75	90	95	99	100
Genre de ménage - ménages unifamiliaux								
. Sans "autres personnes"	0.22	0.05	0.20	0.27	0.33	0.36	0.40	0.56
. Avec "autres personnes"	0.50	0.20	0.48	0.61	0.72	0.74	0.79	0.90
Genre de ménage - ménages non familiaux	0.00	-	-	-	-	-	-	-
Genre de ménage - autres	1.12	0.54	1.05	1.26	1.40	1.51	1.67	1.91
Autres caractéristiques des ménages et des logements	1.00	-	-	-	-	-	-	-
Caractéristiques de la famille de recensement								
Structure de la famille								
. Époux-épouse	0.20	0.09	0.21	0.26	0.29	0.33	0.36	0.42
. Parent seul de sexe masculin	0.64	0.21	0.62	0.81	0.84	0.91	1.04	1.25
. Parent seul de sexe féminin	0.46	0.19	0.45	0.57	0.65	0.69	0.74	0.91
Genre de famille								
. Famille principale	0.23	0.04	0.24	0.28	0.31	0.34	0.39	0.52
. Famille secondaire	0.90	0.62	0.93	1.15	1.28	1.33	1.40	1.49
Groupes d'âge des enfants à la maison	0.78	0.40	0.70	0.91	0.98	1.09	1.19	1.45
Activité de l'époux, de l'épouse, ou du parent seul								
. Époux, parent seul, époux-épouse actifs	0.40	0.23	0.43	0.50	0.55	0.59	0.71	0.93
. Epouse active	0.61	0.41	0.60	0.68	0.74	0.78	0.82	1.15
. Autres	0.72	0.30	0.68	0.80	0.90	0.99	1.12	1.38
Travail en 1985 de l'époux, de l'épouse, ou du parent seul								
. À travaillé en 1985	0.48	0.11	0.45	0.50	0.54	0.57	0.59	0.63
. N'a pas travaillé en 1985	0.93	0.60	0.90	1.04	1.18	1.26	1.30	1.43
Autres caractéristiques de la famille de recensement	1.00	-	-	-	-	-	-	-

Tableau 9. Facteurs d'ajustement de l'erreur-type au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération (RP) - fin

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP						
		1	50	75	90	95	99	100
Caractéristiques de la famille économique								
Structure de la famille économique								
. Familles époux-épouse	0.29	0.13	0.30	0.36	0.48	0.56	0.68	0.91
. Autres	0.56	0.35	0.50	0.66	0.81	0.90	1.06	1.28
Langue maternelle de la personne repère de la famille: Anglais								
. Terre Neuve, Ile-du-Prince-Édouard, Colombie-Britannique	0.25	0.09	0.20	0.31	0.45	0.66	0.91	1.43
. Québec	0.49	0.25	0.47	0.50	0.69	0.83	1.05	1.53
. Autres provinces	0.18	0.07	0.19	0.22	0.24	0.27	0.31	0.49
. Canada	0.27	-	-	-	-	-	-	-
Langue maternelle de la personne repère de la famille: Français								
. Québec	0.12	0.04	0.13	0.17	0.21	0.29	0.36	0.51
. Autres provinces	0.88	0.30	0.90	1.07	1.21	1.28	1.35	1.69
. Canada	0.40	-	-	-	-	-	-	-
Langue maternelle de la personne repère de la famille: Ni le français, ni l'anglais								
. Terre-Neuve, Nouvelle-Écosse	0.75	0.38	0.74	0.80	0.91	0.99	1.10	1.38
. Autres provinces	0.50	0.21	0.45	0.57	0.82	0.84	0.99	1.47
. Canada	0.56	-	-	-	-	-	-	-
Autres caractéristiques de la famille économique	1.00	-	-	-	-	-	-	-

qu'au niveau de la RP. A moins que la région considérée ne soit plus petite que la province, il faut choisir la colonne "Facteur national ou provincial". Les facteurs d'ajustement pour les provinces sont fournis uniquement lorsqu'ils diffèrent de façon significative du facteur national. Cela s'est produit uniquement dans le cas des caractéristiques relatives à la langue maternelle. Si un facteur d'ajustement est requis pour une estimation se rapportant à une région plus petite qu'une province, les "centiles des facteurs au niveau de la RP" permettront d'obtenir une valeur plus précise. Les centiles donnent l'étendue de tous les facteurs d'ajustement au niveau des RP calculés dans l'étude pour les différentes sous-catégories d'une caractéristique. N% des facteurs d'ajustement au niveau de la RP se situaient en-dessous du N^{e} centile et 100 - N% se situaient au-dessus du N^{e} centile. Par exemple, 90% des facteurs d'ajustement au niveau de la RP étaient au-dessous du 90e centile et 10%, au-dessus. On décidera du centile à utiliser selon que l'on veut une estimation de l'erreur-type plus ou moins prudente. Par exemple, "100" nous donnera une estimation très prudente et "75", une estimation qu'il l'est moins.

Les règles suivantes doivent être appliquées pour calculer l'erreur-type ajustée :

(a) Lorsqu'on détermine l'erreur-type d'une estimation relative aux familles ou aux ménages, il faut utiliser le nombre de familles ou de ménages dans la région, pas le nombre de personnes, pour se reporter à la colonne qui convient dans le tableau 8.

(b) À moins que ce soit clairement spécifié dans le tableau 9, les caractéristiques de la famille de recensement et de la famille économique impliquant l'époux, l'épouse, le parent seul, ou la personne repère de la famille ont les mêmes facteurs d'ajustement que ceux mentionnés pour les caractéristiques de la population. Par exemple, le facteur d'ajustement de la caractéristique "plus haut niveau de scolarité atteint par l'époux, l'épouse, ou le parent seul de la famille de recensement" est le même que celui mentionné pour la caractéristique de la population "plus haut niveau de scolarité atteint".

(c) Pour le classement recoupé de deux ou plusieurs caractéristiques, il faut utiliser le facteur d'ajustement le plus élevé des caractéristiques en question.

(d) Tous les facteurs d'ajustement ont été calculés pour des estimations du nombre de personnes, de ménages, de logements, ou de familles, et non pour, par exemple, des unités comme les dollars. Par exemple, les facteurs d'ajustement pour la caractéristique "revenu du ménage" ont été calculés pour des estimations du nombre de ménages ayant un revenu compris dans différentes catégories telles 5 000 - 14 999 $, 35 000 - 44 999 $, etc., et non pour des estimations tel le revenu moyen du ménage.

L'exemple qui suit illustre la façon de calculer l'erreur-type ajustée. Supposons que l'estimation étudiée est la population immigrante en Ontario. L'estimation du recensement de 1986 pour cette caractéristique était 2 081 200. Le chiffre de population obtenu pour l'Ontario au recensement de 1986 était 9 001 170. Comme aucun des deux chiffres n'est très proche des valeurs données dans le tableau 8, il faut appliquer la formule de la page 37 pour obtenir l'erreur-type non ajustée. Dans ce cas-ci, le résultat serait 2 530. Au tableau 9, on voit que le facteur d'ajustement au niveau provincial pour la caractéristique "immigrant" est 1.12. Il s'ensuit que l'erreur-type ajustée pour cette estimation est 2 530 X 1.12 = 2 834.

L'estimation-échantillon et l'erreur-type correspondante peuvent être utilisées pour construire un intervalle dans lequel on peut s'attendre, avec un degré de confiance donné, à trouver la valeur inconnue d'une caractéristique quelconque de la population. Pour cette étude, on a pris un échantillon parmi un grand nombre possible d'autres échantillons de même taille qu'on aurait pu sélectionner à l'aide du même plan de sondage. Les estimations obtenues à partir des données recueillies auprès des différents échantillons ne seraient pas les mêmes. Si l'on construisait des intervalles qui vont de deux erreurs-types au-dessous de l'estimation à deux erreurs-types au-dessus de cette dernière, alors environ 19 intervalles sur 20 comprendraient la "vraie" valeur qui aurait été obtenue au moyen d'un dénombrement complet. On appelle un tel intervalle un intervalle de confiance à 95% (19/20 = 95%). Toutefois, pour pouvoir garantir un niveau de confiance de 95%, les intervalles doivent être calculés en utilisant la "vraie" erreur-type des estimations-échantillon. Les erreurs-types ajustées, qui sont calculées à partir des tableaux 8 et 9, sont seulement des estimations des "vraies" erreurs-types. A noter cependant que les estimations-échantillon aux niveaux provincial et national devraient être assez proches des vraies erreurs-types pour permettre la construction, avec une précision raisonnable, d'intervalles de confiance à 95%. Au-dessous du niveau de la province, l'erreur-type ajustée ne serait peut-être pas assez précise pour permettre la construction d'intervalles semblables.

En prenant l'erreur-type calculée ci-dessus, on peut obtenir un intervalle de confiance à environ 95% pour le nombre d'immigrants en Ontario en procédant comme suit:
2 081 200 ± 2(2 834) soit 2 081 200 ± 5 668.

IX. CONCLUSION

L'échantillonnage fait maintenant partie intégrante du processus de recensement. Son utilisation permet de faire des économies substantielles et de réduire de façon appréciable le fardeau du répondant, ou encore, d'élargir la portée d'un recensement sans frais supplémentaires. Le prix de ces avantages est une introduction d'une erreur d'échantillonnage dans les estimations obtenues à partir de l'échantillon. Cet effet se fait particulièrement sentir sur les estimations de faible valeur du recensement, qu'il s'agisse des chiffres se rapportant à des catégories peu usuelles au niveau national ou provincial, ou de ceux qui correspondent à des régions géographiques de petite taille. Il convient de souligner que les erreurs de réponse et les erreurs de dépouillement contribuent également à l'erreur totale des estimations du recensement et, là encore, ce sont les estimations de faible valeur qui sont particulièrement sensibles aux effets de ces erreurs non dues à l'échantillonnage. De toute façon, elles n'auraient qu'une fiabilité relative même lors d'un recensement intégral. Pour ce qui est du recensement de 1986, les estimations inférieures ou égales à 50 établies à partir des données-échantillon sont généralement très peu fiables tandis que les estimations entre 50 et 500 tendent à avoir des erreurs-types de plus de 10% de leur valeur.

Les méthodes utilisées pour pondérer les données-échantillon en vue d'obtenir des estimations à l'échelle de la population totale ont été appliquées avec succès et, de façon générale, ont permis d'atteindre les niveaux prévus de cohérence des estimations-échantillon et des chiffres de population. Toutefois, la faible cohérence au niveau des SD s'est avérée quelque peu surprenante et ce, en dépit du fait que les méthodes de pondération n'avaient pas été conçues pour contrôler la cohérence au niveau des SD. Une autre exception notable a été le manque de cohérence relativement aux réponses multiples à la question sur la langue maternelle. Cette situation était apparemment attribuable au fait que les répondants interprétaient la question différemment sur la formule 2B que sur la 2A. Les estimations-échantillon de plusieurs autres caractéristiques présentaient aussi un certain biais. Il a été établi que ce biais avait été causé partiellement mais non entièrement par le traitement des données, et le Contrôle et l'Imputation. Le reste du biais provient d'un ou plusieurs facteurs tels que le biais de non-réponse, la déviation des réponses, la sélection, par les recenseurs, d'un échantillon biaisé, etc. Les méthodes de pondération ont toutefois permis de corriger ce biais pour la plupart des caractéristiques. Les estimations-échantillon qui sont demeurées biaisées après la pondération étaient celles qui se rapportaient aux caractéristiques pour lesquelles les chiffres de population étaient peu élevés.

Enfin, certains changements doivent être apportés aux méthodes de pondération pour le recensement de 1991. Les méthodes d'estimation décrites dans le présent document n'ont été que légèrement modifiées depuis leur création en 1971. Depuis, l'élaboration d'autres techniques de pondération a beaucoup progressé. Les langages de programmation disponibles utilisés pour mettre en application les algorithmes de pondération ont aussi été améliorés. Par conséquent, en prévision du recensement de 1991, on étudie actuellement des solutions de rechange à la MIQ et, selon les résultats obtenus, celles-ci devraient produire des estimations plus précises. En outre, les nouvelles méthodes de pondération sont conçues de façon à améliorer la cohérence des estimations-échantillon et des chiffres de population au niveau des SD. Ces améliorations devraient permettre d'accroître

considérablement la fiabilité des estimations sans pour autant augmenter les coûts ou le fardeau du répondant.

ANNEXE

MATRICES DE PONDÉRATION À CLASSEMENT RECOUPÉ

Tableau A1. Lignes de la matrice de pondération des ménages, recensement de 1986

Ménage	Soutien de ménage		Nombre de personnes dans le ménage	Ligne no.
	Sexe	Âge		
Ménages d'une ou plusieurs familles	Homme	15-24	=2	1
			>2	2
		25-34	=2	3
			>2	4
	Femme	15-34	=2	5
			>2	6
	Homme	35-44	=2	7
			>2	8
		45-54	=2	9
			>2	10
		55-64	=2	11
			>2	12
	Femme	35-64	=2	13
			>2	14
	Homme	≥65	=2	15
			>2	16
	Femme	≥65		17
Ménages non familiaux d'une personne	Homme	15-34		18
	Femme	15-34		19
	Homme	35-64		20
	Femme	35-64		21
	Homme	≥65		22
	Femme	≥65		23
Ménages non familiaux de deux personnes ou plus	Homme			24
	Femme			25

Tableau A2. Colonnes de la matrice de pondération des ménages, recensement de 1986

Mode d'occupation	Genre de logement	Colonne no.
Logements Possédés	Maison individuelle	1
	Autre	2
Logements loués	Appartement	3
	Autre	4

Tableau A3. Lignes de la matrice de pondération des personnes, recensement de 1986

Sexe	État matrimonial	Âge	Ligne no.
Homme	Jamais marié	0 - 4	1
		5 - 9	2
		10 - 14	3
		15 - 19	4
		20 - 24	5
		25 - 44	6
		45 - 64	7
	Déjà marié	15 - 24	8
		25 - 34	9
		35 - 44	10
		45 - 54	11
		55 - 64	12
		≥65	13
Femme	Jamais marié	0 - 4	14
		5 - 9	15
		10 - 14	16
		15 - 19	17
		20 - 24	18
		25 - 44	19
		45 - 64	20
	Déjà marié	15 - 24	21
		25 - 34	22
		35 - 44	23
		45 - 54	24
		55 - 64	25
		≥65	26

Tableau A4. Colonnes de la matrice de pondération des personnes, recensement de 1986

Situation des particuliers dans la famille			Langue maternelle	Colonne no.
Membres de la famille	Époux, famille époux-épouse	Sans enfant	Anglais (A)	1
			Français (F)	2
			Autre	3
		Avec enfants	A	4
			F	5
			Autre	6
	Père, ou mère, famille monoparentale		A	7
			F	8
			Autre	9
	Épouse, famille époux-épouse	Avec enfants < 6 ans	A	10
			F	11
			Autre	12
		Sans enfant < 6 ans	A	13
			F	14
			Autre	15
	Enfants dans les familles	0 - 14 ans	A	16
			F	17
			Autre	18
		≥ 15 ans	A	19
			F	20
			Autre	21
Personnes hors-famille	Personne 1		A	22
			F	23
			Autre	24
	Autres membres du ménage		A	25
			F	26
			Autre	27

RÉFÉRENCES

[1] Béland, Y. Results and Methodology of the 1986 Sampling Variance Study, Statistique Canada, rapport interne, Août, 1990.

[2] Brackstone, G.J. et Hill, C.J. The Estimation of Total Variance in the 1976 Census, Survey Methodology Journal, Vol. 2, No. 2, Statistique Canada, 1976.

[3] Brackstone, G.J. et Rao, J.N.K. An Investigation of Raking Ratio Estimators, Sankhya, Volume 41, Series C, Pt. 2, 1979, pp. 97-114.

[4] Bureau fédéral de la statistique. Huitième recensement du Canada, 1941, Rapport administratif du Statisticien fédéral. Ottawa: Imprimeur du Roi, 1945.

[5] _____. Neuvième recensement du Canada, 1951, vol. XI, Rapport administratif. Ottawa: Imprimeur de la Reine, 1955.

[6] _____. Recensement du Canada de 1961, Revue générale, bulletin 7.2-12, no 99-537 au catalogue. Ottawa: Imprimeur de la Reine, 1970.

[7] _____. Sampling in the Census, S.M.S.03.5, Mars, 1968.

[8] Cochran, W.G. Sampling Techniques (Third Edition). New York, John Wiley & Sons, 1977.

[9] Daoust, P. et Bankier, M. Evaluation of Weighting for the 1986 Census of Population, Statistique Canada, rapport interne, Janvier, 1989.

[10] Daoust, P. 1986 Weighting Area Formation Evaluation Report, Statistique Canada, rapport interne, Septembre, 1987.

[11] _____. Analysis of Large 2A/2B Inconsistency for Multiple Responses to Mother Tongue Variable, Statistique Canada, rapport interne, Mars, 1988.

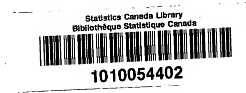
[12] Deming, W.E. et Stephan, F.F. On a Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals are Known, Ann. Math. Statist., 11, 1940, pp. 427-444.

[13] Fellegi, I.P. Response Variance and its Estimation, Journal of the American Statistical Association, 59 (Décembre, 1964), pp. 1016-1041.

[14] Hansen, M.H., Hurwitz, W.N., et Bershad, M.A. Measurement Errors in Censuses and Surveys, Bulletin of the International Statistical Institute, 38 (Août, 1959), pp. 359-374.

[15] Rathwell, S. Results and Methodology of the 1986 Sampling Bias Study and the 1986 2A/2B Consistency Study, Statistique Canada, rapport interne, Août, 1990.

[16] Royce, D. The Use of Sampling in the 1981 Canadian Census, Statistique Canada, rapport interne, Décembre, 1983.



RÉFÉRENCES - fin

[17] Statistique Canada. Recensement du Canada de 1971, Revue générale, vol. VI, partie 1, no 99-740 au catalogue. Ottawa, 1976.

[18] _____. Recensement du Canada de 1976, Rapport administratif, partie 1, no. 99-850 au catalogue. Ottawa, 1980.

[19] _____. Recensement du Canada de 1976, Qualité des données, Série 1: Sources d'erreurs - échantillonnage et pondération, no. 99-844 au catalogue. Ottawa, 1980.

[20] _____. Recensement du Canada de 1981, Guide sommaire: Population totale, no. 99-902 au catalogue. Ottawa, décembre 1983.

[21] _____. Recensement du Canada de 1986, Recensement en bref, no. 99-104F au catalogue. Ottawa, juin 1988.

[22] _____. Recensement du Canada de 1986, Guide à l'intention des utilisateurs sur la qualité des données du Recensement de 1986: Couverture, no. 99-135F au catalogue, 1990.

[23] Stuart, A. Basic Ideas of Scientific Sampling (Second Edition). New York: Hafner Press, 1976.